MINISTRY OF EDUCATION AND TRAINING
VAN LANG UNIVERSITY

ICLIE 2021 VIETNAM

Van Lang University - Faculty of Commerce

PROCEEDINGS OF INTERNATIONAL CONFERENCE ON LOGISTICS AND INDUSTRIAL ENGINEERING 2021

SOCIAL SCIENCES PUBLISHING HOUSE

PROCEEDINGS OF INTERNATIONAL CONFERENCE ON LOGISTICS AND INDUSTRIAL ENGINEERING 2021

VAN LANG UNIVERSITY - FACULTY OF COMMERCE

PROCEEDINGS OF INTERNATIONAL CONFERENCE ON LOGISTICS AND INDUSTRIAL ENGINEERING 2021

SOCIAL SCIENCES PUBLISHING HOUSE

FOREWORDS

The International Conference on Logistics and Industrial Engineering (ICLIE) is an international forum for leading researchers, educators, software developers, and practitioners to discuss current issues, share ideas and experiences on the latest development in the field of Industrial System Engineering, Logistics and Supply Chain Management, and Management Science.

The theme of ICLIE 2021 is: "Industrial Engineering and Supply Chain Management in the Global Business Environment". Nowadays, as the world is becoming heavily integrated, supply chains are becoming more vulnerable to disruptions such as pandemics and natural disasters. Furthermore, the mounting pressure of environmental and socio-economic problems is becoming more apparent. As a result, the aim of the conference is to seek for unfolding new business development opportunities through , Logistics and Supply Chain Management in the global environment.

There are 150 papers submitted to the conference. After being reviewed by experts, 70 papers were accepted to be presented at the conference sections and published in the Proceeding of ICLIE 2021 in English version following in **Interested topics**, but not limited to:

- Sustainable and Green Products
- Design and Performance Optimization of Renewable Energy Systems (Wind, Solar, Wave, Hydro power, etc.)
- Sustainable Industrial Processes
- International Collaborations in Logistics Education and Training.
- Business & Economics.
- Business Ethics
- Business Intelligence
- Business Information Systems
- Business Performance Management
- Business Statistics
- International Commerce
- Smart Logistics and Green Mobility
- Closed Loop Supply Chains
- Green Sourcing and Procurement
- Reverse Manufacturing and Logistics
- Logistics for Healthcare Systems
- Sustainability in Logistics and Supply Chains
- Novel Automation, Information and Data Architectures on the Design of Logistics Systems and Supply Chains
- Intelligent Transportation and Distribution Systems: Theory and Application Models
- Automation and Informatics for Intelligent Transportation Systems
- Port, Supply Chain and Warehouse of The Future
- Distribution Strategies & Packing Management
- Logistics Planning and Control
- Optimization of Logistics & SCM Systems
- Inventory Management

- Decision Support Systems for Logistics & SCM
- E-logistics, E-supply chain, Last Mile Delivery Problems
- IoTs Applications in SCM
- Operation Research/ Optimization
- Modelling, Simulation in Supply chains and Industrial Systems
- Scheduling in Manufacturing & service systems
- Smart City
- And other related topics in the field of Business and Commerce, Industrial Engineering, Management Science, Logistics, and Supply Chain Management

The ICLIE 2021 is held in Ho Chi Minh City, Vietnam; hosted by **Van Lang University (VLU) in collaboration with Ho Chi Minh City University of Technology (HCMUT).**

The Organization Committe appriciates the enthusiastic participation of researchers, policy makers, industry experts, practitioners, and students from all over the country. In addition, we also would like to give a big thanks to the contribution of VLU and HCMUT to the success of the conference.

Again, the The ICLIE 2021 gratefully acknowledges the support from all of you.

On behalf of the Organization Commitee of the International Conference on Logistics and Industrial Engineering 2021.

Conference Organizing Committee

No.	ID	Paper name	Authors	Page
1	4	Application of simulated annealing to optimize the multi-depot vehicle routing problem with time windows	Khai-Kien Lao and Ngoc-Hien Do	1
2	5	Apply statistical characteristics of quanlitative variable to improve productivity and quality: a case study	Ly Duc Minh, Nguyen Quang Sang, Do Ngoc Hien	7
3	6	Apply exploratory data analysis tools into productivity improvement. Case study of nc process at abc company	Ly Duc Minh, Do Ngoc Hien	16
4	7	Application of data envelopment analysis in logistics outsourcing: the case in vietnam	Ngoc-Ai-Thy Nguyen, Thanh-Tuan Dang, Le-Thanh-Hieu Dang and Chia-Nan Wang	22
5	9	Examining the barriers associated with agile manufacturing implementation: case studies in vietnam	Ton Nguyen Trong Hien	29
6	11	Two-stage stochastic programming for production planning	Phan Nguyen Ky Phuc	35
7	12	Enhancement of efficiency by improving workstation: a case study in garment company	Truong Thanh Tam, Nguyen Kim Ngan and Đo Ngoc Hien	40
8	13	Cross Entropy Method and Genetic Algorithm approaches for the Job Shop Problem	Tran Ngoc Quynh, Do Ngoc Hien and Tran Quoc Cong	47
9	18	Does lean itself gain sustainable competitive advantages to manufacturing?	Ton Nguyen Trong Hien	55
10	20	Application of fault detection method and preventive action to improve solar cell production stage performance	Pham Ngoc Phung, Ho Thi Linh, Pham Anh Trong, Do Ngoc Hien	59
11	21	An optimization model for vehicle routing problem with multiple pickup and multiple delivery with heterogeneous vehicle and time window	Phan Nguyen Ky Phuc and Nguyen Le Phuong Thao	65
12	22	Machine learning in minimizing error for forecasting new covid-19 cases in united states: lstm, bilstm, gru	Dao Vu Truong Son and Nguyen Minh Khoi	70
13	23	An application of data envelopment analysis to evaluate the profitability and marketability efficiency of pharmaceutical companies in vietnam	Xuan-Huynh Nguyen, Quoc-Tuan Nguyen and Thi-Kim-Lien Nguyen.	79
14	25	Optimizing production flow by implementation group technology: a case study in educational equipment shop floor	Trương Thành Tâm, Nguyễn Hoàng Phúc, Đỗ Ngọc Hiền	88
15	27	Integrated farm model for sustainable development in binh thuan province, viet nam	Trang Le Thi Kim, Wu C. Kenny, Phong Ho Thanh and Wang Chia-Nan	96
16	29	Application of machine learning and optimization technique On demand and supply planning	Le Dinh Quoc Huy and Phan Nguyen Ky Phuc	102
17	30	Healthy urban planning application in ho chi minh city, A potential for future development toward sustainability	Vo Minh Phuc, Vo Trong Cang	121
18	33	A truck scheduling problem for multi-crossdocking system with metaheuristics	Duong Bao Tram and Phan Nguyen Ky Phuc	127
19	34	Solving vehicle routing problem of a fmcg company in vietnam using k-nearest neighbor algorithm	Tran Quynh Le, Thanh-Tuan Dang and Chia-Nan Wang	138

No.	ID	Paper name	Authors	Page
20	36	Perishable intermodal multiple product supply chain networks uncentainty analysis by discrete-event simulation approach: a case study of fresh fruit supply chain	Luong-Nhat Nhieu, Kha-Hoang Nguyen and Chia-Nan Wang	144
21	38	Preferential policies in attracting foreign direct investment in viet nam	Le Ngoc Thom	154
22	39	Design and analysis of a microgripper based on triple-stair half bridge-type compliant mechanism	Thi Diem-My Le, Minh Trieu-Nguyen Dang, and Chia–Nan Wang	163
23	41	Redesigning the warehouse's material storage location: A case study of the warehouse of An electric motorcycle assembly company	Thi-Huong Pham, Quoc-Anh Lai, Thi-Thu-Thao Le, Thu-Phuong Nguyen, Huu-Duc Huynh and Khai-Kien Lao	167
24	42	Enhancing the garment company's manufacturing facility layout	Huu Phuoc Dang, Phuc Long Nguyen, Quoc Cuong Truong, Duc Huy Nguyen, Huu Duc Huynh and Khai Kien Lao	174
25	43	Demand planning framework in vietnam fmcg industry: a study on the marketing new 4c concept	Bao Le Gia, and Thanh-Tuan Dang	180
26	44	Opportunities for sustainable fashion: student segment in ho chi minh city	Le Vu Lan Oanh and Dinh Tien Minh	191
27	45	Locating hubs for sapote in chau thanh district, Tien giang province, viet nam	N. V. Chung and H. N. Bao Vi	200
28	47	Medical waste collection vehicle routing problem: a case study in thu duc city	Tong Hong Phuc and Ha Thi Xuan Chi	206
29	48	An application of simulation and optimization to outpatient appointment scheduling systems - a case study of health care clinic in ho chi minh city	Le Dinh Hoang Dang, Nguyen Hoang Huy	212
30	49	The intrinsic entrepreneurial motivation: a review on the decision of students to participate in the project-based competition program	Tran N.H. Ngan and Andraz J.M. and Nguyen H.L. Thuy	222
31	50	Automatic process resume in talent pool by applying natural language processing	Quynh Trinh, Thanh-Tuan Dang	234
32	51	Multi-resource routing and scheduling problem-a case study of home healthcare service in ho chi minh city	Trinh Minh Dang, Nguyen Hoang Huy	241
33	60	Forecasting future performance of vietnam banking industry using grey prediction and data envelopment analysis	Kim-Phong Tran, Chia-Nan Wang and Hector Tibo	253
34	61	Application of machine learning in multi-class classification: a case study of life insurance risk assessment	Khanh Ho and Son Dao	262
35	62	Mathematical optimization model for dispatching ride-hailing	Cao Nguyen Dang Minh and Nguyen Vo Cong Thanh	270
36	63	An improved mathematical model for dynamic pricing In ride-hailing platforms	Cao Nguyen Dang Minh and Doan Le Thao Vy	277
37	65	Using k-nearest neighbor for fall detection A feature selection framework using grey wolf optimizer	Phuong Thao Nguyen and Vu Truong Son Dao	284

No.	ID	Paper name	Authors	Page
38	66	Application of machine learning in credit card default prediction	Nga Vu, Son Dao	293
39	67	Application of simulation in healthcare system: a case study of outpatient clinic in a hospital in ho chi minh city	Nguyen Truong Nguyen, Doan Le Thao Vy, Le Quang Trung, Nguyen Vo Cong Thanh and Nguyen Bui Thanh Truc	302
40	69	Minimize the travel distance of the pickers in the warehouse	Dang Vi Anh, Duong Vo Nhi Anh, Nguyen Hoang Huy	316
41	70	Design supply chain network for reverse logistics scancom case study	Vo Hong Thanh Nhi, Anh Duong	326
42	71	Design supply chain network for pangasius feed to the mekong delta - woosung vina company case study	Nguyen Le Cat Quynh, Duong Vo Nhi Anh	338
43	73	Optimal storage allocation for products in warehouse A case study of ych protrade company for suntory pepsico's product	Nguyen Le Binh Minh, Duong Vo Nhi Anh, Ngo Thi Thao Uyen	348
44	74	Mathematical modelling in transportation method Application for rice of can tho city	Nhat Nguyen, Anh Duong, Ly Tran and Uyen Ngo	366
45	75	Split delivery vehicle routing problem with time windows	Nguyen Le Thanh Thuy, Duong Vo Nhi Anh, Do Vinh Truc, Tran Van Ly	374
46	76	Mechanical properties of specimens 3d printed with virgin and recycled acrylonitrile-butadiene-styrene and polyethylene terephthalate	Thi Thanh Tam Nguyen, Bernd Muschard, Chia-Nan Wang	382
47	78	Quality management system in garment industry – case of thuan phuong group	Pham Cao Van, Bui Dinh Tan and Vo Trong Cang	393
48	80	Biodegradable substances prediction in drug discovery Using random forest	Van T.H. Nguyen, Son V.T. Dao	401
49	81	The study and design of crm system in admissions counseling and support for students: case of international college of ho chi minh city	Nguyen Quynh Lam, Do Doan Trang, Vo Trong Cang Tang Minh Huong, Nguyen Thi My Le	411
50	82	Innovative improvement of sample order tool for internal system - case of scancom company	Nguyen Quynh Phuong, Tran Thi Mong Tuyen and Vo Trong Cang	418
51	83	Application of system engineering in design of solar energy system: a case study	Dat Minh Le, Phuong-Nhi Thieu Pham, Quang-Hao Duong, Nam-Anh Diep	426
52	85	An empty container optimization in distribution network of a logistics company under uncertainty - the case of asia	Hoang Thi Bich Ngoc	445
53	88	Dynamic integration of process planning and scheduling with due date assignment	Le Thuc Dan Trinh Nguyen Hoang Huy, Duong Vo Nhi Anh	451
54	89	Development of services providers for ho chi minh city become a logistics center in vietnam - objectives and solutions	Tran Chanh Nghia, Mai Thi Dieu, Nguyen Chi Khang, Nguyen Minh Luat, Dinh Thi Hoai Nhi, Do Doan Trang, Vo Trong Cang	461

No.	ID	Paper name	Authors	Page
55	93	Supplier Evaluation and Selection Process for Textile Supply Chains in Vietnam: an MCDM Approach	Viet Tinh Nguyen, To Tam Ngo, Ngoc Nhat Linh Le and Thi My Tran Nguyen	466
56	94	Distribution Centre location selection for FMCG supply chains: a MCDM approach	Van Thanh Nguyen, Thi Kim Huong Nguyen, Thai Thanh Bui and Thanh Sang Thai	472
57	95	Supplier evaluation and selection process for renewable energy projects: an mcdm approach	Syed Tam Husain, Tin Giang Cam, Minh Thu Nguyen and Hong Vy Vo	478
58	96	Location selection process for geothermal energy projects in vietnam: an mcdm approach	Viet Tinh Nguyen, Phi Truong Cao, Thi Tuyet Mai Truong and Thi Ngoc Vu	483
59	98	Real estate business in vietnam and green marketing	Nguyen Minh Ngoc and Nguyen Hoang Tien	489
60	99	The effect of e-health-related knowledge, privacy & security concerns on the intention to use e-health – a patient's perspective	Doan, Thi Lien Huong and Tran, Dinh Long	497
61	100	Application of capacitated vehicle routing problem with multi-depot and multi-trip: a case study of wine retailer in vietnam	Le Tran Phuong Anh, Nguyen Hoang Huy, Do Vinh Truc	505
62	101	Real estate business in the face of changes in customer behavioral psychology during the covid-19 pandemic in vietnam	Nguyen Minh Ngoc, Nguyen Thi Hai Yen, Nguyen Hoang Tien	515
63	103	Mathematical programming formulation	Ngo Hoang Minh Anh, Anh Duong, Huy Nguyen and Ly Tran	524
64	104	Enhancement of sales forecast process for a jewelry retailer using sarima and winters triple exponential smoothing	Thuy Duong Nguyen, Ngoc Hien Do	533
65	107	Heterogeneous fleet vehicle routing problem with time window: a case study in dksh vietnam	Nguyen Van Chung, Ngo Thanh Phuong Anh and Luong-Nhat Nhieu	539
66	108	Order picking optimization mathematical model with order batching approach: a case study of vietnam third-party logistics company	Nguyen Van Chung, Le Nguyen Tung and Luong-Nhat Nhieu	546
67	110	Difficulties and problems in establishing and applying iso 9001:2015 standard at enterprises in vietnam	Ho Thi Phuong Thao and Le Huong Duong	553
68	112	Applying lean in a garment manufacturing company: a case study	Nguyen Huu Phuc and Trinh Thi Thuy An	560
69	113	A study of inland waterways logistics in southern vietnam	Vo Trong Cang, Cao Viet Hieu and Nguyen Anh Tuan	565
70	114	Research and improvement on warehouse operation process based on discrete-event simulation model	Nguyen Thi Phuong Linh, Nguyen Huu Tho and Do Ngoc Hien	572
71	14	Work design at steel manufacturing to improve safety of preventive maintenance	Luong Thanh Tri, Nguyen Tan Thanh, Nguyen Anh Kiet, Nguyen Phuong Nguyen, Do Ngoc Hien	582

APPLICATION OF SIMULATED ANNEALING TO OPTIMIZE THE MULTI-DEPOT VEHICLE ROUTING PROBLEM WITH TIME WINDOWS

Khai-Kien Lao[1] and Ngoc-Hien Do[2]

[1] HCMUT, Industrial Systems Engineering Dept., Ho Chi Minh City, Vietnam,

luminousarceye@gmail.com

[2] HCMUT, Industrial Systems Engineering Dept., Ho Chi Minh City, Vietnam,

hienise97@hcmut.edu.vn

Abstract. Competitiveness of last mile delivery companies can be judged through customer service, service policies, etc. Yet, quality and price usually are most concerned by users. Hence, optimizing these two factors is a need in any transport company. This research applies the MDVRPTW model to the routing of a delivery company. More specifically, the main subject is the delivery service from depots to customers. Using simulated annealing algorithm, the research objective is to optimize delivery cost while minimize failed deliveries in order to raise the company's competitiveness. Additionally, the model is cross checked with OR-tools to test its validity and performance. As a result, the model has reduced delivery cost around 15% compared to the current deployed method of the company while considering all time windows of recipients.

Keywords: simulated algorithm; vehicle routing problem; multi-depot vehicle routing problem with time window; optimization

1. Introduction

With online shopping trend on the rise, E-commerce has become one of the markets with highest growth rate. As a result, the demand for logistics in this market is huge and both its investment and development opportunity are great. However, the VECOM E-Business Index Report (2018) shows that there are still many obstacles for a breakthrough in the coming period. Less than half of the interviewees claim to be satisfied with online shopping mainly due to goods quality and service response level. This means there is still a huge potential customer base out there waiting for businesses to conquer them. Regularly, customer judge a product or service by many criteria. However, the two outstanding are price and quality. This is more true when it comes to delivery services since the two criteria also are what firms compete over. From the viewpoint of customers, these companies rival through their price. But, nowadays, they actually focus on reducing operating cost. Cutting down travel cost is one of the most common strategy used. Relating to this matter, VRP is a mathematical model studied with the purpose of planning the route to go through all required locations to achieve one or many objectives while considering system constraints.

The vehicle routing problem with time windows (VRPTW) is a problem where the customers need to be serviced in a time interval, called time window. The work on VRPTW can be divided into two categories: exact optimization and heuristic algorithms. Larsen (1999) and Chabrier (2006) acquired improvements in Solomon's benchmark problems. Cordeau et al. (2001) solved these problems using tabu search algorithm while and Gambardella et al. (1999) tried to solve it with genetic algorithm.
The multi-depot vehicle routing problem (MDVRP) is another variant of VRP where the starting location of vehicles are different. According to Elin Haerani (2017) In the process of optimizing, multiple depot vehicle routing problem has three stages: grouping, routing, and scheduling stages. With the emergence of problems with multiple starting points, the location-routing problem (LRP) survey was carried out by Drexl and Schneider (2014 & 2015), Prodhon and Prins (2015). Recently, research on using heuristic algorithms to solve the above problem has also appeared quite a lot: searching for a local optimal solution by Subramanian (2013), tabu search algorithm by Cordeau and Maischberger (2012), Escobar (2014), genetic algorithm by Vidal (2014), hierarchical hybrid meta-heuristic by Shimizu and Sakaguchi (2014).

Appear later and more complicated, the combination of the two problems mentioned above is the multi-depot VRP with time window (MDVRPTW). Due to being a NP-hard problem, it is usually solved by suitable and good enough heuristic algorithms rather than exact method. A recent study in this field was by Heechul Bae, Ilkyeong Moon (2016). These two authors' research aims to meet the service level while ensuring transportation costs.

This paper aims to apply the MDVRPTW in a delivery company to optimize travel cost while improve its service quality by getting all delivery on time. Since exact method for this problem can only be done with small instances, this paper will solve the problem through simulated annealing algorithm, a heuristic approach. The result will then be cross-checked with OR-tools developed by Google and compared to the method currently used by the company.

2. Problem description

The research object of this paper is a delivery service involving a fleet of vehicles, a few depots and a set of customers need to be served at certain time interval. Currently, the goods is grouped according to their destination district so that they are forwarded to corresponding depot. The routing is then done separately for each depot and the trips are assigned to driver for delivery.

However, the method mentioned is not sufficient enough. Through collecting and analyzing data about this service, though the fulfilled demands is high, nearly 90%, only 79.74% of the customers are successfully served at their time window. The remaining customers are served at another time which increase unnecessary travel cost. This is due to the company only aim at minimize travel cost but not consider enough for time window which lead to most of the failed deliveries. Additionally, the grouping being used now does not ensure the customers are grouped to the nearest depot.

Therefore, in this paper, vehicle routes are formed in a way that:

- The total travel cost is minimized
- All demands are fulfilled in their time window
- Each customer is served exactly once by a vehicle
- The load of a route does not exceed vehicle capacity
- Each route starts and ends at assigned depot
- Duration of each route must not exceed the preset limit

3. Mathematical model formulation

The mathematical model for the problem is a 2-stage process: group first, route second. The group technique used is to assign the customers to the nearest depot instead of only based on which district they are in. The objective is to minimize total travel cost, which is proportional to travel distance, will then turn into multiple sub model VRPTW for each depot. Though the final solution found may not be globally optimal but should be good enough as local optimization. Figure 1 illustrates the solution main idea.

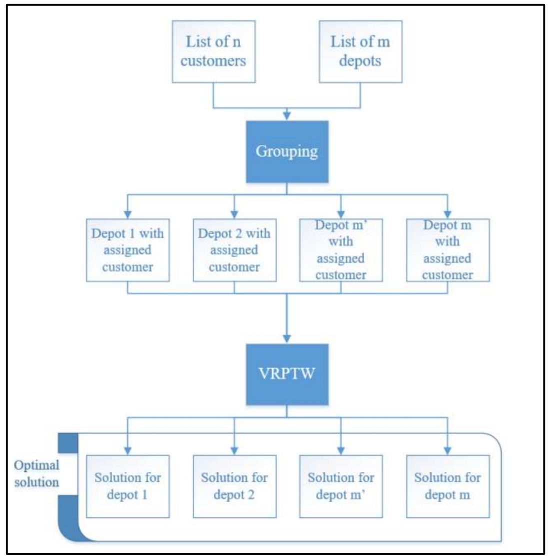


Figure 1. The 2-stage MDVRPTW

The sets, parameters and decision variables are described below:

Sets

C Set of customers $\boldsymbol{C} = \{1, 2,\ldots, n\}$ where n is total number of customer

$C(d)$ Set of customers assigned to depot $\boldsymbol{d}$

D Set of depots $\boldsymbol{D} = \{1, 2,\ldots, m\}$ where m is total number of depot

V Set of customers and depots $\boldsymbol{V} = \{\boldsymbol{C} \cup \boldsymbol{D}\}$

A Set of movable arc set $\boldsymbol{A} = \{(\boldsymbol{i, j}) \mid (\boldsymbol{i, j}) \in \boldsymbol{V}\} \setminus \{(\boldsymbol{i, j}) \mid (\boldsymbol{i, j}) \in \boldsymbol{D}\}$

K A fleet of vehicles $\boldsymbol{K} = \{\boldsymbol{k_1, k_2,\ldots,k_L}\}$ with identical capacity $\boldsymbol{Q}$

$K(d)$ A fleet of vehicles assigned to depot $\boldsymbol{d}$

Parameter

q_i Demand of customer $\boldsymbol{i}$ for $\boldsymbol{i} \in \boldsymbol{C}$

e_i Opening time of node $\boldsymbol{i}$ for $\boldsymbol{i} \in \boldsymbol{V}$

l_i Closing time of node $\boldsymbol{i}$ for $\boldsymbol{i} \in \boldsymbol{V}$

c_{ij} Travel distance from customer $\boldsymbol{i}$ to customer $\boldsymbol{j}$, $\forall(i, j) \in \boldsymbol{A}$. The distance is determined by using *vincenty* package in Python

t_{ij} Travel time from customer $\boldsymbol{i}$ to customer $\boldsymbol{j}$ plus service time at customer $\boldsymbol{j}$, $\forall(i, j) \in \boldsymbol{A}$. The travel component is calculated by using corresponding travel distance divide with average moving speed in the city. The service component is a constant.

Decision variables

For each depot ***d*** and assigned customer set ***C(d)***, a fleet of vehicle ***K(d), V(d) = C(d)*** ∪ {***d***}. Consider *0* is the index for depot in ***V(d).*** The model has two sets of variables

x_{ijk} — x_{ijk} = 1 if vehicle ***k*** travel arc (***i, j***), zero otherwise, ∀(i, j) ∈ V(d), k ∈ K(d)

s_{ik} — A non-negative time variable record the time when vehicle ***k*** serve customer ***i***. If setting departure time at depot $e_0 = 0$ then $s_{0k} = 0$

$s_{jk} = s_{ik} + t_{ij}$ if vehicle ***k*** arrives at customer ***j*** and serve it between its time window right after customer ***i***

$s_{jk} = e_j$ if vehicle ***k*** arrives at customer ***j*** before its time window

If vehicle ***k*** does not serve customer ***i*** then s_{ik} is null

With declared notations above, the detailed mathematical model can be stated as:

Objective

$$min \sum_{k\in K(d)} \sum_{i\in V(d)} \sum_{j\in V(d)} c_{ij} x_{ijk} \qquad (1)$$

Constraints

$$\sum_{k\in K(d)} \sum_{j\in V(d)} x_{ijk} = 1, \forall i \in V(d) \qquad (2)$$

$$\sum_{i\in C(d)} q_i \sum_{j\in V(d)} x_{ijk} \le Q, \forall k \in K(d) \qquad (3)$$

$$\sum_{j\in V(d)} x_{0jk} = 1, \forall k \in K(d) \qquad (4)$$

$$\sum_{i\in V(d)} x_{ihk} - \sum_{j\in V(d)} x_{hjk} = 0 \qquad (5)$$
$$, \forall h \in C(d), \forall k \in K(d)$$

$$\sum_{i\in V(d)} x_{i0k} = 1, \forall k \in K(d) \qquad (6)$$

$$s_{ik} + t_{ij} - K(1 - x_{ijk}) \le s_{jk} \qquad (7)$$
$$, \forall i, j \in V(d), \forall k \in K(d)$$

$$e_i \le s_{ik} \le l_i, \forall i \in V(d), \forall \in K(d) \qquad (8)$$

$$x_{ijk} \in \{0,1\}, \forall i, j \in V(d), \forall k \in K(d) \qquad (9)$$

Expression meaning:

- (1) express the objective which is to minimize travel distance or travel cost of delivery from depot *d*
- (2) and (5) indicate that each customer is served exactly once by a vehicle and that vehicle must leave after every visit
- (3) ensure vehicle won't exceed its capacity at every stop
- (4) and (6) indicate each vehicle departs and finishes a trip at the same depot
- (7) and (8) force vehicles to serve customers in their time windows
- (9) set x_{ijk} as binary variable

4. Simulated Annealing algorithm

Simulated annealing (SA) is a probability search algorithm, which is an optimization method that can be applied to find the global optimization of the goal function and avoid the local optimization by accept a worse solution with a temperature dependent probability *β*.

Simulated annealing is of mechanical origin, working on the principle of physical metallurgy. In physical metallurgy the metal is heated to a high temperature and cooled slowly so that it crystallizes in a low energy configuration. If cooling does not occur slowly, the solid that does not reach the low-energy configuration will freeze to an unstable state (locally optimal structure).

SA uses a global control variable, the temperature variable *β*. Initially β_0 is at a very high value and then gradually decreases. During the SA search, replace the current solution by randomly choosing a neighbor solution with a probability that depends on the difference between the objective function value and the control parameter *β*.

The optimization process is continued until a global minimum is found or the total number of moves exceeds a predetermined maximum number of moves or a specific criterion is met. The transition at one temperature ends when thermal equilibrium is reached. After reaching thermal equilibrium, the temperature is reduced lower. If the system is not frozen and no global minimum is found, the loop continues. The system freezes when T approaches the user-specified final T temperature.

The pseudo code below illustrates SA process.

#Simulated Annealing
n ← *Max loop*
X ← *Initial solution*
f (X) ← *Goal value function of solution X*
β_0 ← *Initial temperature*
$\beta = \gamma(\beta_0)$ ← *Function of temperature*
best = X
now = X
loop = 0
While *loop ≤ n*:
    From *now*, generate neighbor solutions

Select best neighbor solution *X'* and compare to current best solution *X*

If $f(X') < f(now)$:

$now = X'$

If $f(X') < f(best)$:

best = X'

else:

If $e^{(f(X') - f(now))/\beta} > r$ *# r is a random number from Uniform distribution (0,1)*

$now = X'$

loop = *loop* + 1

Continue to next loop

Return *best* when max loop reached

4.1. Initial solution

To apply SA, initial solution is an important input since it will affect the searching process and quality of the final solution. This paper creates initial solution for each set of depot and customers based on Greedy algorithm. The main idea is to find and add the nearest node from current position that satisfy all constraints to form a route. The pseudo code below will demonstrate this process.

Set **D** = [0]

Solution **S** = [ ]

Goal value **Z** = 0

k = 1

For vehicle k:

Set timer **T** = 0

Goal **G** = 0

Route **P** = [ ]

Current position **i** = 0

Step 1:

If |**P**| = Max capacity, go to Step 4

Else determine all customers can be served:

$\mathbf{F} = \{ j \notin \mathbf{D} \mid \mathbf{T} + t_{ij} \leq l_j \ \& \ \mathbf{T} + t_{ij} + t_{j0} \leq l_0 \}$

If none customer available go to Step 4, else continue to Step 2

Step 2:

Determine if any customer from **F** are ready to be serviced on arrival:

$\mathbf{F'} = \{ j \in \mathbf{F} \mid \mathbf{T} + t_{ij} \geq e_j \}$

If exist, select the one closet to current position **i** and back to Step 1

$\mathbf{T} = \mathbf{T} + t_{ij}$ $\mathbf{P} \leftarrow j$ $\mathbf{D} \leftarrow j$

Else go to Step 3

Step 3:

Choose the next ready customer in F

$\mathbf{T} = e_j$ $\mathbf{P} \leftarrow j$ $\mathbf{D} \leftarrow j$

Back to Step 1

Step 4:

Calculate **G** base on sequences from **P**

Z = **Z** + **G**

Add **P** into **S**

Continue to next vehicle k = k + 1

4.2. Neighbor solutions

The most commonly known method to generate neighbor solutions is exchanging nodes in a route or nodes of two or many routes. The method used in this paper is the later one which is to select one customer each from a chosen pair of route and swap while considering time aspect. One thing to note is this method will only be used if the current solution having two or more route. The process of generating neighbor solutions goes like this:

Step 1: From current solution, pair all route to be parent route used from step 2. For example, if the solution have 3 routes {a, b, c}, there are 3 pairs (a,b), (a,c) and (b,c). If the solution have 4 routes, there are 6 pairs.

Step 2: Select a random customer **x** from one route. Calculate *time window overlap* between it and every customer from the other route. Select the customer with highest overlap, called customer **y**.

Step 3: If the overlap between **x** and **y** is positive, exchange position of **x** and **y.** Otherwise, delete either **x** or **y** from the one route and insert it into the remain route. Figure.2 illustrates the exchange process.

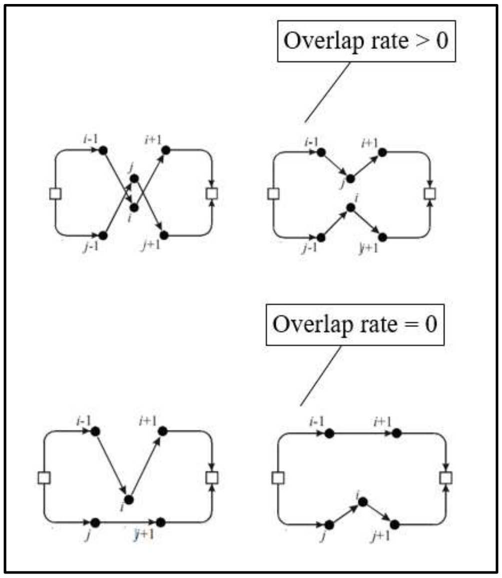


Figure 2. Exchange process

Step 4: Verify if the new routes satisfy all constraints to add it to the neighbor solution pool. Back to step 1 if there are pair to exchange. If not, the process ends here.

5. Results

The proposed model and algorithm are coded in Python language and implemented on Google Colaboratory. The input data used to test the model is of a normal working day queried from company database. The program is run several times with many scenarios of max loop such as 50,100, 250, 500, 1000 and 1500 to collect data for analysis. The parameters are as follows. There are 5 depots, 1262 customers, a very large fleet of vehicles with the which can serve a maximum of 20 customers, the preset time limit for all vehicle is 300 minutes since the routing is complete. As mentioned, the method will be compared to current method and OR-Tools. Table 1 and Figure 3 present total travel distance (in meter) of solutions for the data set.

Table 1. Results of MDVRPTW-SA

Max loop	Initial solution	MDVRP TW - SA	OR-Tools	Current method
50	2134	1934.4	1818	2191.25
100		1908.6		
250		1890.6		
500		1873.6		
1000		1860.4		
1500		1852.6		

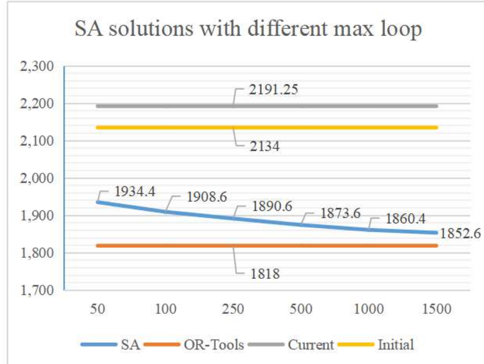


Figure 3. Goal value of different solutions

The results imply:

- Using Greedy empirical algorithm as the initial solution, the program produces an objective function value with an improvement of about 2% compared to the original distance value and an improvement of 11- 15% after the SA algorithm. This shows that the algorithm, although simple, still gives quite good results.
- Regard to the SA algorithm, in the first 1000 iterations, we can see the improvement of goal value is quite distinct and quickly as the number of iterations increase. However, the improvement is not significant after that, so maybe 1000 iterations with a run time of about 400 seconds is a good parameter for this data set.
- The proposed thesis model and the model using OR-tools, respectively, improve about 15% and 17% compared to current routing method. This also means the company can reduce shipping costs by more than 15% per delivery while minimizing delivery failures which can greatly cut the cost for company operation.
- The best solution that the model this paper proposed can give up to now is approximately 98.96% effective compared to OR-tools, i.e the error is approximately 2%. According to research by Razali (2011), this error level is acceptable.

Therefore, the model has been tested and proven to be effective, so it can be put into experimental evaluation. However, there is still room for improvement. As mentioned, it can be seen that the OR-tools model gives quite good results, so it is possible that if we take this result as an initial solution instead of the empirical algorithms, then proceed to improve it with SA algorithm, the model will give better results than that. After calibrate the program and run it, the result is truly better than the Greedy then SA. With 1000 loops, the model results in a total of 1773m travel distance which improve about 19% with the running time of 324 seconds.

However, consideration should be given to applying the model using OR-tools because the model may give inaccurate results. The reason is that OR-tools is currently only developed to handle integer data, but not decimals, irrational numbers, etc. The results received here are also the results when the program has rounded up the distance between moving points.

The research has gone through the steps of idea generation, matching with reality based on actual data and capable of meeting the company's improvement expectations. Next we will summarizes the results of the study as well as discusses aspects and directions for further research and improvement.

6. Conclusions

The paper propose an application of simulated annealing algorithm, a heuristic approach, to the MDVRPTW of a real case study. The results show that proposed optimization model does reduce around 15% travel cost in the routing phase of the operation. The study also shows that we absolutely can apply the optimization model MDVRPTW which was cross checked through comparison with tools available in the market. The model has taken into account the time frame of receiving goods from the customer, so it will partly help in minimizing delivery failures.

7. Reference

Chabrier A. (2006), Vehicle Routing Problem with Elementary Shortest Path based Column Generation. *Computers and Operations Research*, **33(10)**, 2972-2990.

Cordeau, J.F., Laporte, G., Mercier, A. (2001), A Unified Tabu Search for Vehicle Routing Problem with Time Windows, *Journal of the Operational Research Society 52*: 928-936.

Cordeau, J., and Maischberger, M. (2012), A parallel iterated tabu search heuristic for vehicle routing problems, *Computers and Operations Research,* **39**, 2033- 2050.

Drexl, M., and Schneider, M. (2014), A survey of the standard location-routing problem. *Working Paper LPIS-03/2014*, *Logistics Planning and Information Systems*, *TU Darmstadt,* Germany

Drexl, M., and Schneider, M. (2015), A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research*, **241**, 283-308.

Elin Haerani, Luh Kesuma Wardhani, Dian Kumala Putri, Husni Teja Sukmana (2017), Optimization of Multiple Depot Vehicle Routing Problem (MDVRP) on Perishable Product Distribution by Using Genetic Algorithm and Fuzzy Logic Controller (FLC).

Escobar, J. W., Linfati, R., Toth, P., and Baldoquin, M. G. (2014), A hybrid granular tabu search algorithm for the multi-depot vehicle routing problem. *Journal of Heuristics*, **20**, 483-509.

Gambardella, L. M., Taillard, E., Agazzi, G.," (1999) MACS-VRPTW: A Multiple Ant Colony System for Vehicle Routing Problems with Time Windows", *New Ideas in Optimization,* McGraw-Hill

Heechul Bae, Ilkyeong Moon, Multi-depot (2016), Vehicle Routing Problem with Time Windows considering Delivery and Installation Vehicles, Applied Mathematical Modelling

Larsen J. (1999) Parallelization of the vehicle routing problem with time windows. *Ph.D. Thesis IMM-PHD-1999-62*, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Prodhon, C., and Prins, C. (2014), A survey of recent research on location-routing problems. E*uropean Journal of Operational Research*, **238**, 1-17.

Razali, N. M., & Geraghty, J. (2011, July) *Genetic algorithm performance with different selection strategies in solving TSP* (**Vol. 2**, pp. 1134-1139)

Shimizu, Y., and Sakaguchi, T. (2014), A hierarchical hybrid meta-heuristic approach to coping with large practical multi-depot VRP. *Industrial Engineering and Management Systems*, **13**, 163-171.

Subramanian, A., Uchoa, E., and Ochi, L.S. (2013), A hybrid algorithm for a class of vehicle routing problems. *Computers and Operations Research*, **40**, 2519 - 2531.

Vidal, T., Crainic, T.G., Gendreau, M., and Prins, C. (2014), Implicit depot assignments and rotations in vehicle routing heuristics. *European Journal of Operational Research*, **237**, 15-28.

Vietnam E-Business Index 2018 Report

APPLY STATISTICAL CHARACTERISTICS OF QUANLITATIVE VARIABLE TO IMPROVE PRODUCTIVITY AND QUALITY: A CASE STUDY

Ly Duc Minh[1*], Nguyen Quang Sang[2], Do Ngoc Hien[3]

[1] Faculty of Commerce, Van Lang University,Ho Chi Minh City,minh.ld@vlu.edu.vn

[2] Faculty Automobile of Technology, Van Lang University, Ho Chi Minh city,sang.nq@vlu.edu.vn

[3] Industrial Engineering Major, Ho Chi Minh University of Technology, hienise97@hcmut.edu.vn

Abstract. Lean Six Sigma represents a management approach for driving innovating processes inside a company in order to achieve superior results. It involves a practical analysis based on facts, aiming the innovation and growth, not only the efficiency of processes. Quality is more than making things without errors. It is about making a product or service meet the individual perception of a customer about the quality or value. Therefore, in what regards Lean Six Sigma, the concern is not only to "do the things right" but also to "do the right things right". We focus on the impact of implementing the Lean Six Sigma approach on companies, seeking for what changes and benefits it brings. The key elements it aims at are achieving the best quality, the lowest cost, getting the shortest lead-time, stressing on waste elimination. The requirements of a company for its implementation and the strategy to obtain the maximum practical outcome are investigated. Furthermore, we conduct a comparison analysis with the other methods of the total quality management and see why Lean Six Sigma is a more desirable approach

Keywords: Continuous Improvement, DMAIC, Lean Six Sigma, Productivity and Quality improvement, PDCA, 7QC tools, Mechanical production line

1. Introduction

Over the past several decades, there has been a growing awareness of the need for quality improvement in the industrial sector. B. Mrugalska et al (2018) proposed implementing the DMAIC (Define-Measurement-Analysis-Improvement-Control) cycle as an element of continuous improvement in the machining process, to increase productivity and process efficiency. submit. Besides, in order to meet the needs of customers and respond to the demands of the competitive market, manufacturing enterprises should consider implementing activities to reduce waste products. P. Thakur et al. (2021) proposed building and using tools such as pareto charts, voice of customers, quality control story to collect data, analyze and observe the causes of waste generation. Products. I.L. Nunes (2015) proved that lean six sigma is one of the tools to help companies continuously improve, enhance their strong competitiveness, and achieve high efficiency in manufacturing enterprises. as well as services.

Besides, M.M. Uddin (2020) thinks and executes using basic quality tools and methods like Define, Measurement, Analysis, Improvement, Control (DMAIC) for root cause analysis on waste generation by using 5 whys, checklists, and implementing employee awareness training more effectively. To meet the increasing demand of special orders from many customers, from many different companies, to improve production lines by cutting cycle time without high value for the products. machining cost function. M.A.L.E. Salleh et al (2018) propose to apply Lean Six sigma tools such as DMAIC to continuous improvement of the production process to improve product quality, process output, reduce waste, reduce waste and reduce costs in production. To improve the performance of machining machines in manufacturing plants, B. Phrukasaphanrat (2019) suggests applying six techniques in six sigma such as DMAIC and innovative tools to improve the manufacturing process of machining machines. Along with that, the training, new design of machine tool components, machine jigs and data analysis, improvement of the documentation system, modification of some machine components, improvement of a number of methods. method at the processing machine to improve machine efficiency.

This paper focuses on the application of Lean Six sigma, quality control story and improved tools in data analysis and improvement of some jigs at machining machines to eliminate waste, reduce processing time, reduce waste, increase machine efficiency. The next chapters will implement the content. Chapter 2 discusses the method of applying the DMAIC cycle in practice to improve processes. Chapter 3 presents the results obtained and discusses the application of the results to similar processes.

2. Methodology

2.1 Definition

As Fig 1, Liner Guide Products have 5 components. Motor Bracket, Rail and Bearing Bracket stand still. Ball Screw are operated to be the same as circular. Block move from Motor Bracket to Bearing Bracket and opposite. Block is moved elements. So that, Quality of Block focus for Linear Guide products.

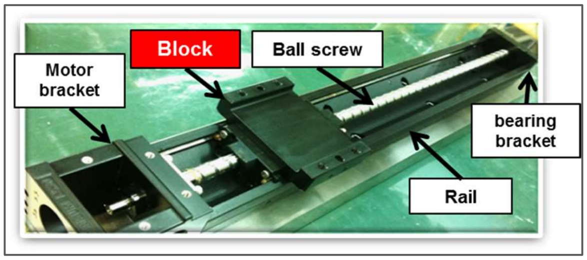


Figure 1: Liner Guide Products

Confirm shipment quantity of Linear Guide products from 7.2020 to 10.2020 Productivity of products size 30 is shipped very much to be compare with the others size.

Figure 2: Shipment quantity

As a high-volume product line, it is necessary to satisfy customer requirements. Check Block's unsatisfactory product data from the assembly stage and find high block error rate products. As Fig 3, Product Block error wide accounts for 73%.

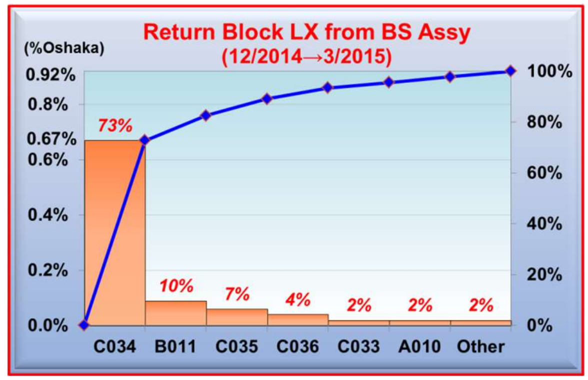


Figure 3: Defect ratio

C034: Block wide, B011: Block is not good slide C033: Block roll, C035: Block NG A008:Scratch, C036: Block floss Verifying the product of the wide block error according to the product line, found that the rate of product size 30 generated errors accounted for 57% as Fig 4.

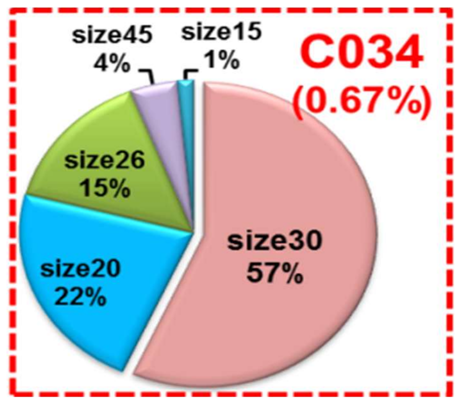


Figure 4: Defect ratio base on product size

As Fig 5, Block Wide error has 3 types: Deflection of the ball hole and threaded hole position.. The block pattern of the threaded groove is not good. Shaft's lace groove is not good. In order to perform an action to improve the wide block error for the Linear Guide product line, we chose to improve the deflection error and choose size 30 as the research object.

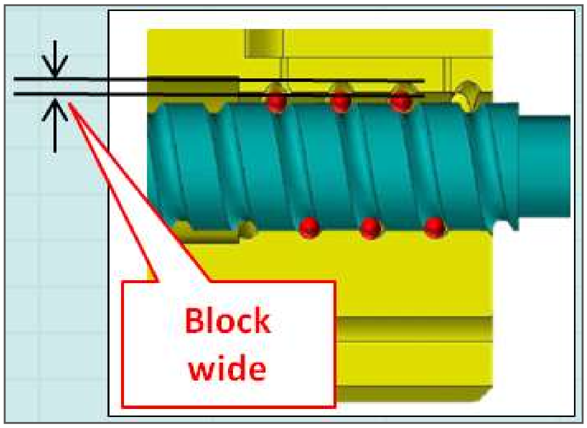


Figure 5: Block wide

We decided to set up goals fell wide block error rate from 57% to 0% for size 30. The objective is clear, proceed to analyze the current situation, causes and take action to improve.

2.2 Measurement

Analyzing the current status of Block machining process and confirming the process of machining Tap is the Tap process and the lapping process. (Fig. 6).

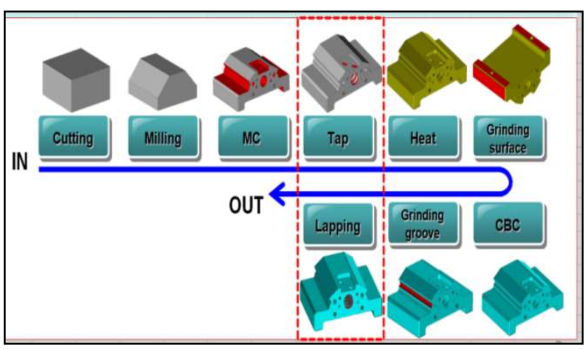


Figure 6: Process flow chart

The failure error is not achieved due to the failure of the tap hole in the machining way. We conduct analysis of tap hole machining steps for block products. Step 1: Convenient threaded groove. Step 2: machining the return hole. (Fig 7)

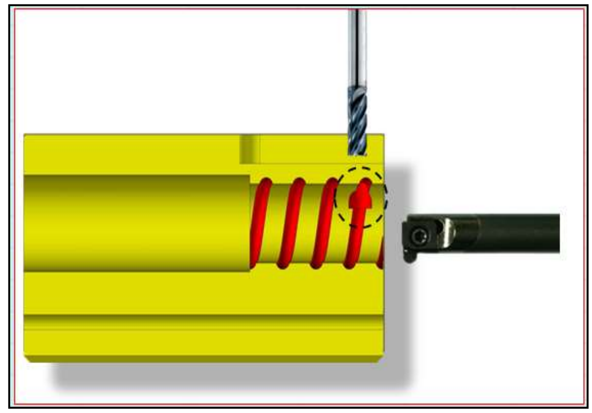


Figure 7: Processing step

while turning thread holes and ball bearings make position difference. The reason is that it cannot be precisely machined because the dimensions cannot be measured. (Fig 7). To check the bore position deviation by using the magnifying glass (Fig 8) to visually determine the deviation size and adjust the machining tolerances manually into the NC machine program.

Figure 8: Inside Measure tools

It is hard to see the deviation because of the small error. Completely based on the judgment of the operator, the machining results are inaccurate because the size cannot be measured accurately. The position difference of the ball hole affects the trajectory of the ball causing a phenomenon. (Fig 9).

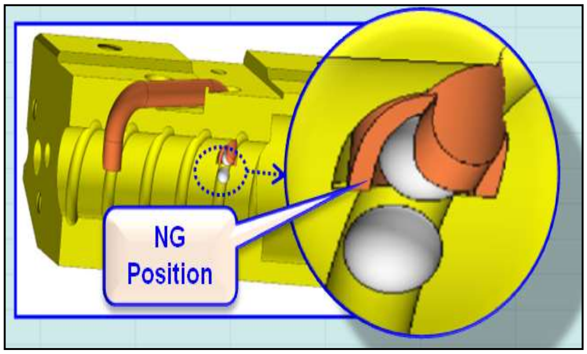


Figure 9: Defect position

Position deviations affect product quality in the process of using balls that do not move well. (Fig 10). In order to check the error of the bouncing hole, we have to disclose the product cutting work and measure the bore hole size,việc cắt sản phẩm làm phát sinh phế phẩm chiếm tỹ lệ 2.38%.

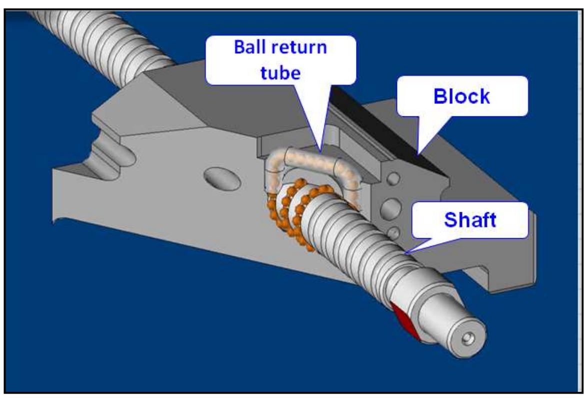


Figure 10: Structure of products

Time for cutting and measuring takes 36.5 minutes. (Fig 11) .

No.	Step Name	Time (min)	Process ○	Move ●	Inspect ◊	Wait ▲
1	MC machine test first sample	5	○			
2	Move to cut machine	1.5		●		
3	Cut sample	4	○			
4	Move to measure machine	2		●		
5	Measure sample	19.5			◊	
6	Calculation of measurement data	1.5			◊	
7	Come back to MC machine	2.5		●		
8	Setting MC machine program	0.5			◊	
	Total	36.5				

Figure 11: Time chart

After checking the dimension on the sample with the magnifier, the MC operator adjusted the hole size tolerance on the MC machine. (Fig 12)

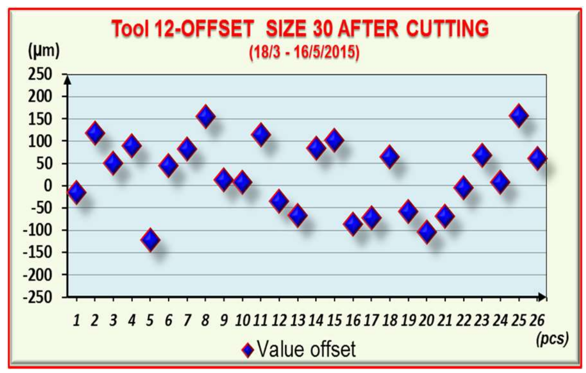


Figure 12: Value offset

A waste of $ 1,352 / year for waste cutting materials and waste processing time is $ 940 / year. The accuracy level is low and when a large test deviation has to be cut again to check again. Not safe when cutting blocks. Use a sensor measuring system that connects a computer to measure hole deflection after machining and save measurement data in a computer. The results are determined through the measurement system on the computer screen and divided into 6 levels from A, B, C +, C-, D and E. (Fig 13)

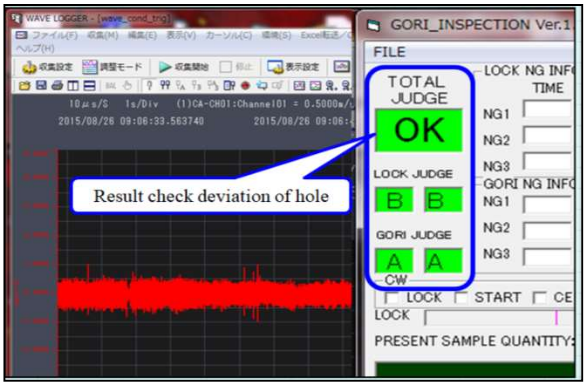


Figure 13: Inspection system

Type E rate accounts for 2.5%. Type D ratio accounts for 16.6%. Position deviations affect deviations test results. Set goals as follows.

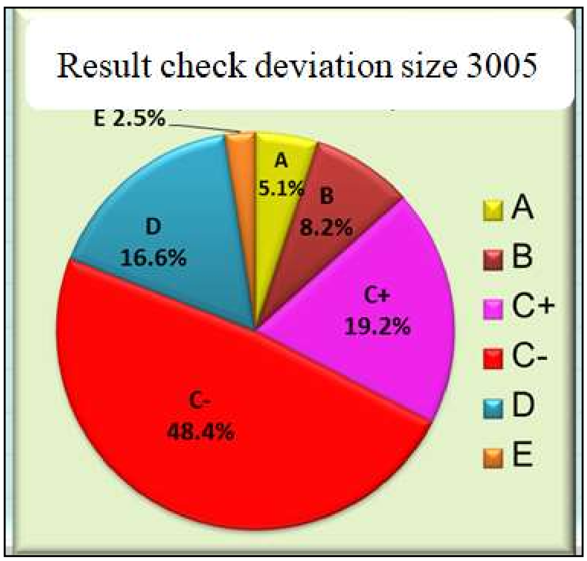


Figure 14: Deviation of each size product

1.Reduce 25% product ratio due to error C034 (Wide Block) size30 (from 0.38% to 0.1%).

2.Reduce the rate of error waste products for size 3005 products to "Zero".

3.Reducing 50% of product grade D rate for products with size 3005 (from 16.6% to 8.3%) because Type D is at risk of becoming Type E.

4.The rate of products damaged by cutting head parts at MC size 30 to "Zero" (cutting rate: 0%).

2.3 Analysis

To find the cause of a hole deviation error, I use the fish bone chart to analyze it and as a result identify 3 causes of a hole deviation error. (Fig 15)

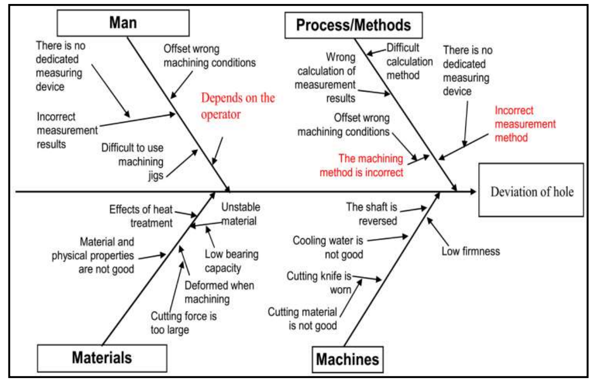
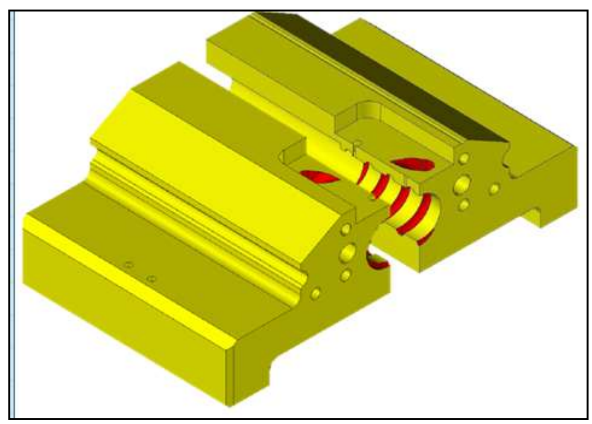


Figure 15: Fishbone analysis defect

To verify the cause caused by the wrong measurement method, I cut the product out and used the magnification meter to measure 4 sizes, 1, 2, 3 and 4. Then calculate the size required by MC machine residual processing. (Fig 16)



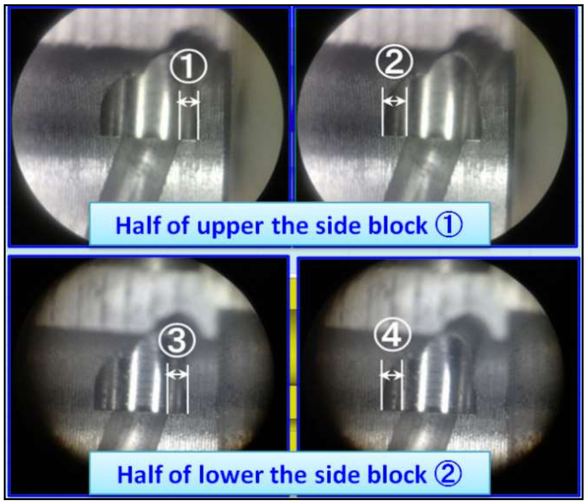


Figure 16: Dimension for offset

To calculate the hole size deviation, I follow the formula to calculate the amount of deviation X:

X = ((①+④)-(②+③)) ÷ 2

And

Quantity Offset = X / 2

The employee inspects the size by eye so the error due to incorrect capture is generated. In addition, multi-point measurement has a high probability of deviation and the risk of calculating the size is wrong.

Analyzing the causes of errors and losses related to depends on the operator. Using correlation chart to verify the relationship between employee's machining experience and hole size measurement results. (Fig. 17).

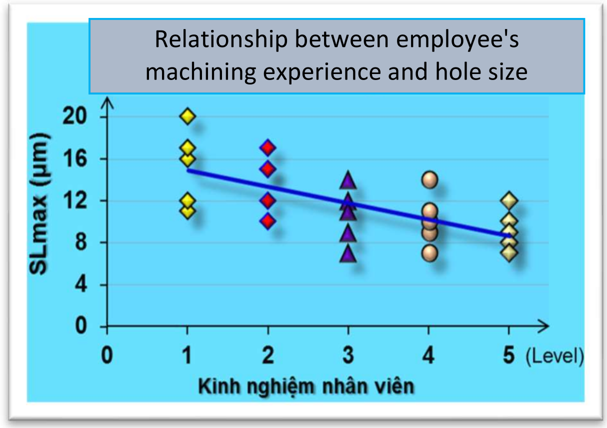


Figure 17: Relationship chart

There is a high correlation between employee experience and size accuracy. The more experienced the employee is, the more accurate the measurement results are. To verify the machining method is incorrect. I installed each of the different blocks with 1 rail and 1 fixed shaft, and then tracked the test results.

Test samples: Type : LXW3005, Rail : Rail master, Shaft : Shaft master, Block : WB 3005 (10 pcs).

The deviation error test results in variations between different blocks (Fig 18).

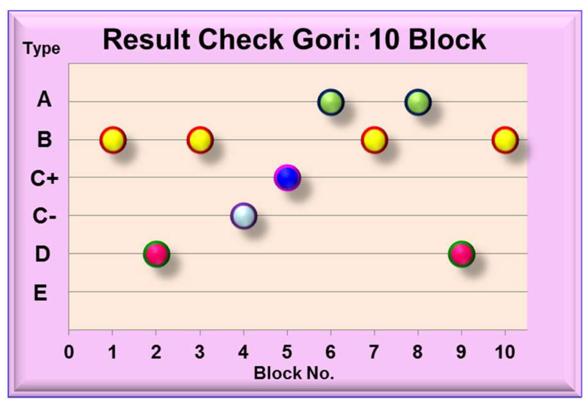


Figure 18: Check Gori

Different deviations between threaded and ball holes cause different GORI test results. Processing size according to current method is not stable

2.4 Improvement

To improve errors due to inaccurate measurement methods, I built the contact measuring instrument. I proceeded to design 3D models for contact tools. (Fig 19)

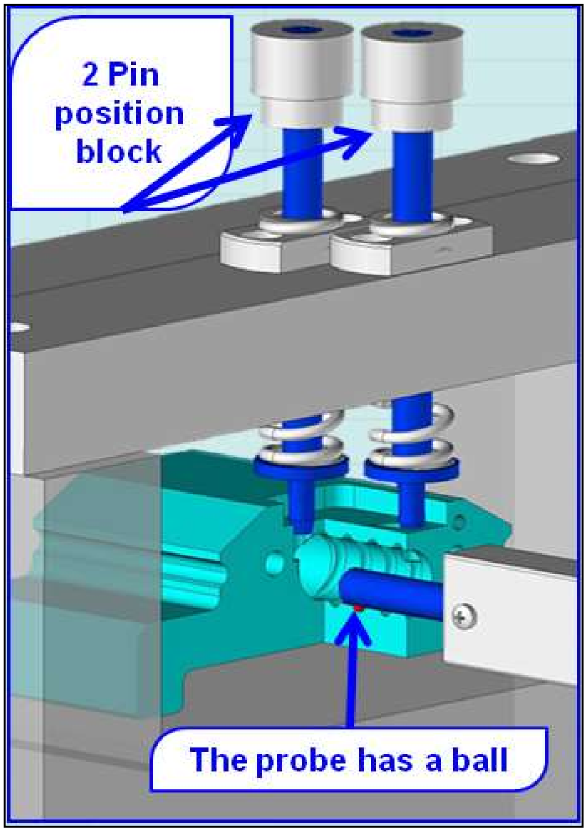


Figure 19: Improvement jigs

The cost for a jig is: 840 USD, including 2 main parts: Manufacturing cost: 200 USD and Meter: 640 USD.

The characteristics of the jig is to use a Stage set as the pin positioning and pin clamping mechanism is fixed so the operation is simple. (Fig 20)

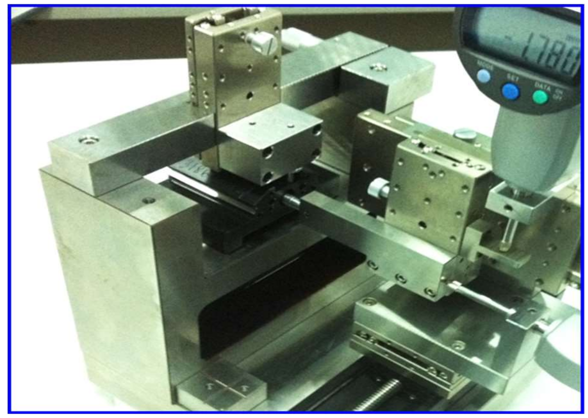


Figure 20: Actual improvement jigs

To verify the stability of the measuring instrument. I conducted the measurement as follows: using 1 product and measuring 10 times continuously, using a chart to confirm the deviation of the measurement results. Result: deviation is 0.67 micron. (Fig 21)

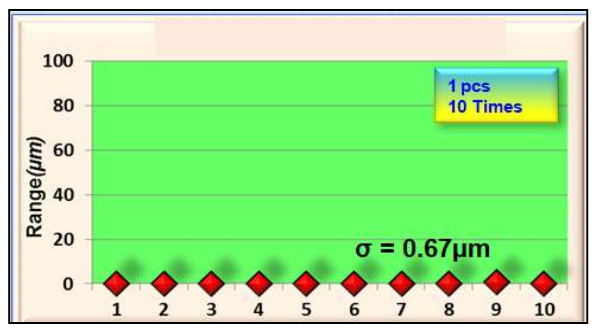


Figure 21: Range of dimension

The result after using contact jig is direct measurement of ball return hole and ball groove, accurate and stable measurement results, simple use, no experience, no need to cut products.

After applying the measuring jig to the production process and re-checking the operation time as well as the operation at the MC stage, it is significantly reduced. Before applying the contact measuring jig, the operation time is 36.5 minutes, after applying the contact gauge, the operation time is 6 minutes. Total reduction time is 30.5 minutes (equivalent to 84% reduction). Carry out a check on the correlation between the employee experience and the measurement of the size by correlating the graph and the result is not affected by the employee's experience after using contact jigs.

No need to cut so the time decreases. Precision jigs. No dependence on experience. Check the rate of error arising after using jigs to contact the MC stage. E-waste products were returned to Zero, no longer repaired, or discarded. Type D ratio has dropped to 7.6%. (Fig 22)

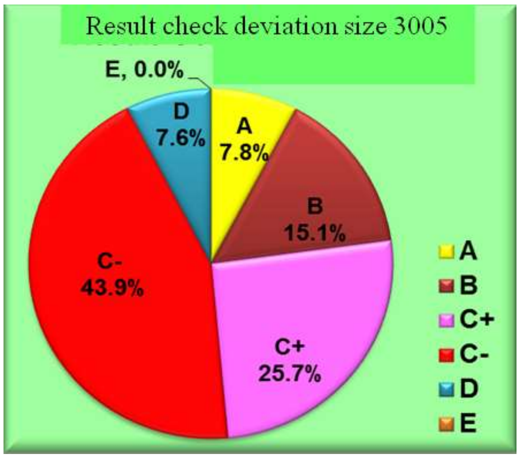


Figure 22: Deviation after improvement

2.5 Control

In order to control measurement data not through human recognition in the test sheet, to prevent errors when recording the measurement results by hand. I proceeded to connect the tool kit to the online data processing system via the USB data transfer port. At the same time, integrating histograms into the online data processing screen so that we can visually observe the stage capacity through the Cpk index. Furthermore, in order to realize when to offset the excess of the machining on the MC machine, I have integrated the chart control chart into the computer screen and the staff knows the point and size needed to offset the residual processing amount. for MC machine, If the data case is out of specifications through control chart. (Fig. 23).

Figure 23: Control chart and histogram online

3. Results and Conclusion

3.1 Results

The result in terms of quality is the precise machining product. No more products out of specifications. As a result, in terms of processing capacity, the MC is the capacity of turning the threaded hole to be guaranteed. The results in terms of improving the operation at the stage of MC machining are simple measurement operation, less errors. Training time for staff decreased from 1 week to 30 minutes.

The results in terms of productivity improvement at the MC processing stage are the first detailed test time: reduce 313h / year ($ 940 / year) Processing costs: reduce 201 Pcs / year ($ 3,472 / year) and Time to repair goods due to: reduce 308h / year ($ 926 / year). The results in terms of reducing materials at the MC processing stage are the number of damaged products: reducing 201 Pcs / year ($ 2407 / year). In addition, to ease the deployment of jigs exposed to machining lines and to be operated well, actions should be taken such as: Training for staff how to use new measuring tools. Calibration periodically (every 6 months). Use the master check and master master regularly. Prepare check sheet check daily jigs.

3.2 Conclusion

I have a better understanding of the importance of precise measurement of machining dimensions. If it is not possible to measure the size correctly, there is no way to accurately process. Measured the size at the invisible position. And it has been "seen" to be dimensionally transformed through stages that have not been done so far to help control output quality in the best way. At first everyone thought it would be better to buy a better measuring device, but those measuring instruments were only measured with images at a very high cost. We have changed the way we think and have built accurate measuring jig with a low cost of 840 USD. And we will continue to improve the technical strength of the engineering team to create inexpensive but high-precision fixtures in the company.

4. Reference

Beata Mrugalska et al (2018), Using Six Sigma DMAIC to improve the quality of the production process: a case study, *Procedia Social and Behavioral Sciences* 238 (2018) 590 – 596, doi:10.1016/j.sbspro.2018.04.039

Pankaj Thakur et al (2021), Analysis and optimization of properties of paint materials for reduction of paint defects in agro products, *Second International Coference on Aspects of Materials Science and Engineering (ICAMSE 2021)*, doi.org/10.1016/j.matpr.2021.02.349

Isabel L. Nunes (2015), Intergration of Ergonomics and Lean Six Sigma. A model proposal, *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015*, doi: 10.1016/j.promfg.2015.07.124

Mohammad Moin Uddin (2020), Improving product quality and production yield in wood flooring manufacturing using basic quality tools, *International Jounal for Quality Research 15(1) 155-170*, ISSN 1800 – 6450, DOI-10.24874/IJQR15.01-09

N. A. M. Salleh et al (2018), Cycle Time Improvement by Lean Six Sigma for the Cutting Tool Production Line, *Journal of Mechenical Engineering*, Vol SI 5(1), 240-264, 2018, ISSN 1823-5514

B. Phruksaphanrat et al (2019), Six sigma DMAIC for machine efficiency improvement in a carpet factory, Songklanakarin J.Sci.Technol, 41 (4), 887-898, Jul-Aug.

H. Nishimura, (2015), Quality control and Qualityi mprovement, JMAM, ISBN 978-4-534-05320-6

Ritsushi Tsukuda, (2012). A Practical guide of Toyota Production System, JMAM, ISBN :978-4-8207-4788-8.

Harukato, (2012), The Fundamentals of Production Managament, JMAM, ISBN 4-8207-4069-5.

APPLY EXPLORATORY DATA ANALYSIS TOOLS INTO PRODUCTIVITY IMPROVEMENT. CASE STUDY OF NC PROCESS AT ABC COMPANY

Ly Duc Minh[1], Do Ngoc Hien[2]

[1] *Faculty of Commerce, Van Lang University, Ho Chi Minh City, minh.ld@vlu.edu.vn*

[2] *Industrial Engineering Major, Ho Chi Minh University of Technology,*

hienise97@hcmut.edu.vn

Abstract: Mold industries continue to face the challenges of improving productivity, in order to remain successful, in a highly competitive industry. As year on year, every input cost goes higher, it is important to provide better products at lower costs to the end consumer by improving the efficiency of the plant. Today the company has a production capacity of almost 52399 pcs per day. This capacity cannot be utilized properly due to various losses. The project aims at identifying the root cause of these losses and to propose methods to increase the productivity. DMAIC technique is used for productivity increase with why - why analysis in the analysis stage to find out the root causes of losses. Pareto chart is used to plot various factors that affect productivity. This project mainly focuses on reasons for power failure and absenteeism, which is a threat to the organization. In order to improve the productivity of the organization, suggestions were made to manage the power failure and absenteeism, thus a better system was proposed for the organization under study.

Keywords: DMAIC Analysis; Lean Six Sigma; PDCA; 7QC tool

1. Introduction

In today's competitive world, planning productivity is one of the main components of successful industrial organizations. Increase in productivity can reduce the cost of work and raw material on the production unit or an increase in output. Productivity depends on various factors like labour, machinery, capital, temperature, raw materials, quality etc. So, each and every factor has its own contribution. The focus of this study, productivity of yarn, has a great influence in Mold industry. In this environment, company managers must tackle various problems regarding men and machines. In the present scenario the operating costs are increasing, and the sales are decreasing, while customers have turned out to be more selective and demanding. Mold industry managers must thus consider how to maintain profitability in a declining market, while providing high quality products to increasingly sophisticated customers. Firms can concentrate on different methods to improve the productivity and operational efficiency of the system.

2. Literature review

Today, with the rapid development of Industry 4.0 and the penetration of technology into the DMAIC (Define, Measurement, Analysis, Improvement, Control) method, Lean manufacturing creates enhanced productivity and continuous improvement. Industrial production processes. P. A. Vargas et al (2021) proposes the application of lean tools such as lean, visual management and standardization in process improvement to reduce machining time, optimize and minimize waste. Tools in Six Sigma (DMAIC) is a guide to problem solving and process improvement and product improvement. L. Girmanava et al (2017) researched using tools such as flowcharts, charts, pareto charts, P-FMEA, (Process-Failure mode and effect analysis), cause-and-effect diagrams, and logical analysis. control the outsourcing process in the enterprise. A. N. S. NEJIB et al (2019) uses a combination of the PDCA cycle (Plan, Do, Check, Action) with the DMAIC cycle to realize continuous improvement and control each step of execution in DMAIC. To meet the increasing demand for orders from many customers, from many different customer companies, to improve production lines by reducing machining time without high value for cost.

NAM Salled et al (2018) proposes to apply Lean Six sigma tools such as DMAIC to continuous improvement of the production process, to improve the output quality of products, reduce waste, reduce waste and reduce waste. costs in production. Sumadi et al (2021) proposed the application of DMAIC techniques to process control and improvement, quality improvement and waste reduction at the production stage by applying tools such as pareto charts, bone diagrams. fish, to analyze the causes of defective products and propose P-FMEA methods to control quality and the impact of defective products on production cost savings. To improve the performance of machining machines in the manufacturing plant. B.

Phruksaphanrat et al (2019) suggested the application of 6 DMAIC techniques in Six Sigma and improvement tools to improve the factory's production process, along with the combination of training and new design of components. in the machining machine, analyze the data and improve the documentation system, modify some components and improve some methods at the machine to improve machine performance. A. Shahin et al (2010) proposed the application of poka yoke in troubleshooting at the design and manufacturing stages of the process.

3. Problem definition

The production capacity of the plant is 165.439 pcs per month. But the actual production rate is only 52.399 pcs per month. Thus the plant has a loss around 113.040 pcs per month, which in all together causes a huge loss in terms of productivity and thereby in profit ratio. From the study conducted at the company, it is evident that the production line is almost perfect but there exists process method losses in Mold lines. The ultimate objective of this project is to conduct a detailed study of various factors reducing productivity and increasing productivity by reducing various losses at the critical processes. Also it aims to find out different problems facing employees and impact of working environments on employee performance.

4. Research methodology

The project aims at eliminating various losses causing low yield or productivity. Various factors of men and machines are to be observed and analysed. In this project DMAIC technique is used. It includes an overview of process, current situation analysis, scope for improvement targeting, implementation of countermeasures. Pareto chart and why- why analysis is used to get an outline of the various root causes of losses in these sectors. Cost estimation and payback analysis are used to analyse the cost and payback.

5. Data analysis and discussion

5.1. Definition

There are many new workers to apply Mold line and Train to use machine in Line not enough skill. Therefor, out put line volunm is low productivity. Other sites, happen Defect ratio is high. Especialy, Out put at NC Latch process is low produtivity to compare with Other machine line. It is Bottle neck process of Mold Line and must make action to improvement productivity for NC Latch process due to time limited within 4 months. (Fig. 1)

Theme	Impact	Emergency	Ability	Total	Estimate term
1	5	5	3	75	6 months
2	2	5	3	30	5 months
3	**2**	**5**	**5**	**50**	**4 months**

Figure 1: Time limited

5.2 Measurement

There are 6 process in Mold line and NC Latch Process is the last process. (Fig 2)

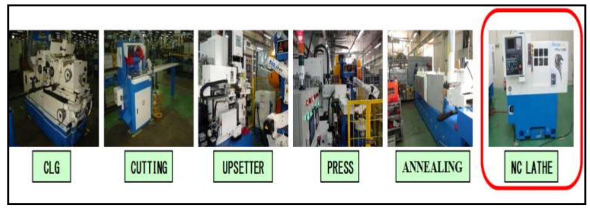


Figure 2: Process flow

Out put of the last process are out put of Line. Therefor, NC Latch process are the most important. Having 2 steps process at NC Latch process,It is call: NC1 and NC2. And Ouput of NC1 are 52399 pcs per month and Output of NC2 are 56619 pcs per month. But Output of CLG are 165439 pcs per month. (Fig. 3)

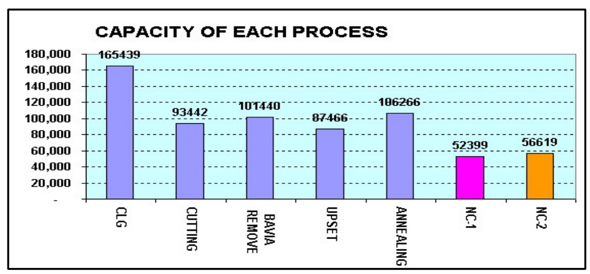


Figure 3: Capcity of each process

So, the next step is to identify the critical factors that reduce the productivity and to analyse how to reduce these losses and thus to improve productivity for NC Latch process

5.3. Analysis

Main responsibility of NC Latch are process at Heat of product. (Fig 4)

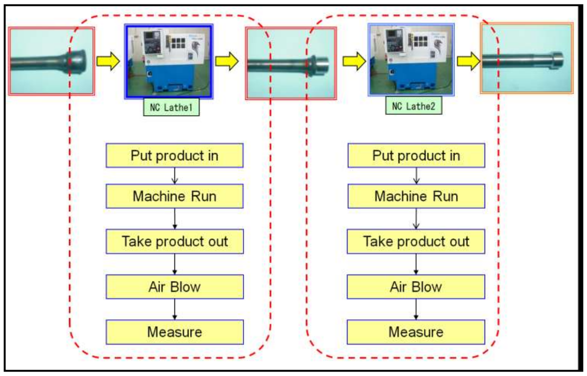


Figure 4: Process NC Latch

Using Video action for NC Latch and Anlaysis time between Man – Machine of NC1 process and NC2 process and Determine 2 items of Handle time and Checking time must be to make action for reducing. (Fig. 5)

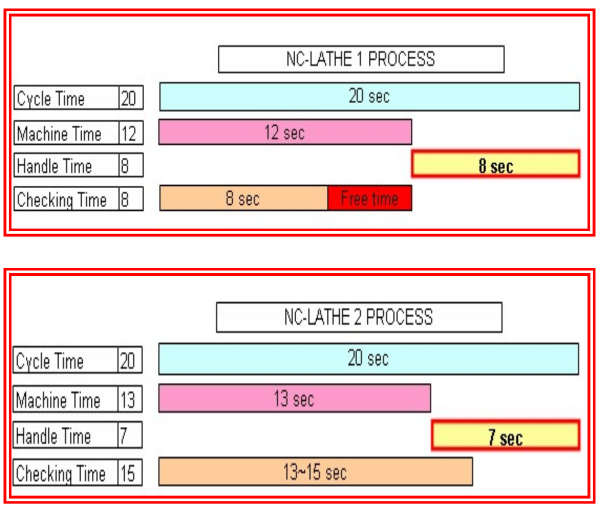


Figure 5: Cycle time of NC Latch process

Affinity Diagram Analysis as major reasons for loss of productivity was identified as power failure and worker absenteeism, Affinity Diagram analysis was conducted on these problems to find out the real causes regarding these problems. (Fig 6)

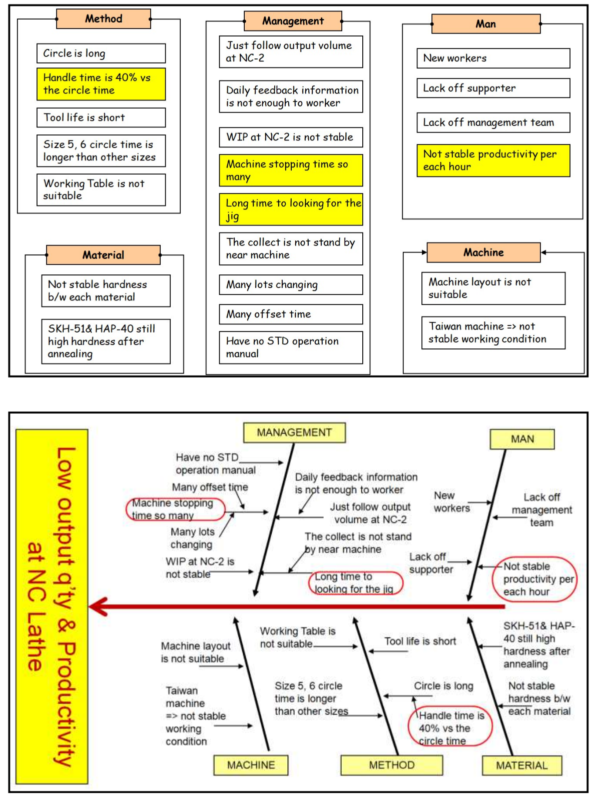


Figure 6: Fishbone Diagram

About Method: Handle time is 40% via the circle time. About Management: Machine stopping time so many and Long time to looking for the jig. About Man: Not stable productivity per each hour. Man-Machine Chart: As waste action for loss of productivity. Affinity Man – Machine Chart was conducted on these problems to find out the real causes regarding these problems. (Fig 7)

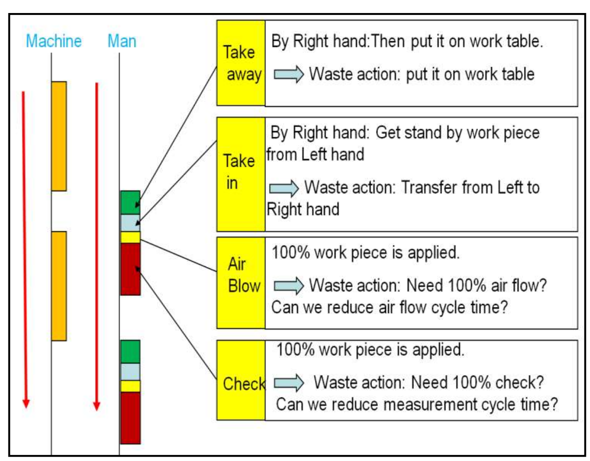


Figure 7: Man – Machine chart time

There have some waste actions on Handle time. It need to be kaizen for NC Latch process. Waste action are include: Take away, Take in, Air flow and Check. by Pareto chart analysis, we can say that more than 60% of total loss of productivity are due to look for jig and stopping time. Back step shortage is occurred only for 2 months. It is because of the sudden change of product and industry has their own reason for that.

We can neglect productivity loss due to Maintenance and repair, other reasons as it accounts for only a small percentage. So we can say that in the Mold industry the main factors that contributes to lower productivity is looked for jig and stopping time. (Fig. 8)

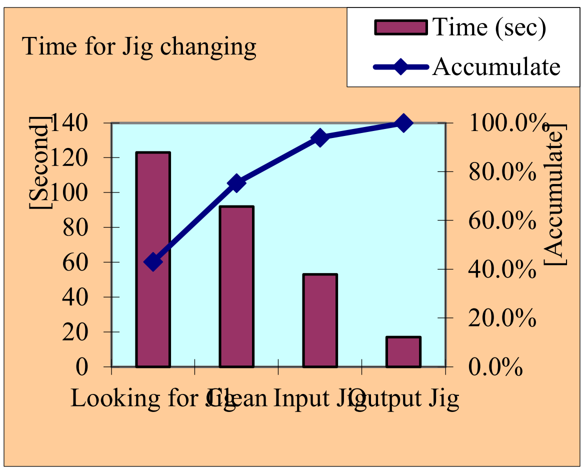


Figure 8: Pareto chart

For Improvement productivity for NC Latch process, must action to improvement 3 points: Reduce handle time, Reduce stopping time and Kepp stable productivity.

5.4. Improvement and Control

The purpose of this step is to identify, testing and implementation of alternative solutions to the problem, in part or in whole. Identify feasible solutions to reduce the key root causes with the aim of fix and prevent process problems. Here the root causes of men and machine are improved. Improvement is mainly done by:

a) Suggestion of a new productivity gain by improvement operation method.

Change use manual air gun into auto air gun and fix place. Result of this action is reduced from 3 sec to 2 sec. (Fig 9)

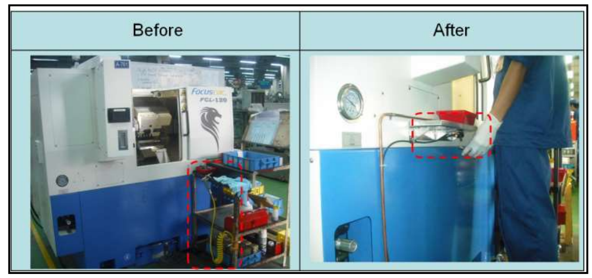


Figure 9: NCL Kaizen

b) Suggestions to improve the working environment and motivation of workers.

For prevent lose time by moving to get jig, tools for setting products. To make layout to put jig, tools place near NC Latch machine and fix place to put each jig, tools to be clear. (Fig 10)

Figure 10: Outside NCL Process

By this action, we can reduce collect changing time from 304 sec to 210 sec.

c) Suggestion of a new measurement method and measurement tool layout.

Change measurement tools from Caliper with setting free into Micrometer with fixing into the jig near NC Latch machine. (Fig 11)

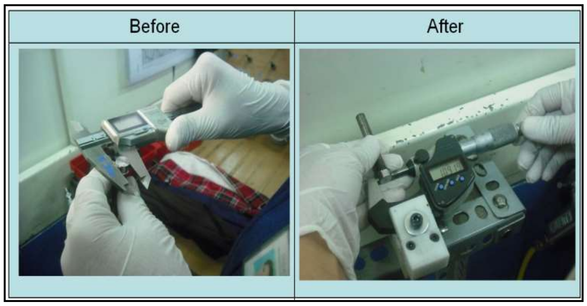


Figure 11: Measure at NCL Process

By this action, we can reduce measurement time from 6 sec to 4 sec.

d) Visual management control method for output daily, weekly and monthly.

For control status output of daily, weekly and monthly products by visual. Set monitoring screem and put in line Mold where NC Latch process can see clearly information in screem. (Fig. 12)

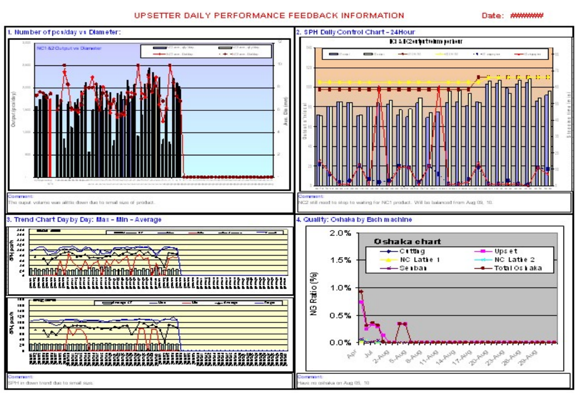
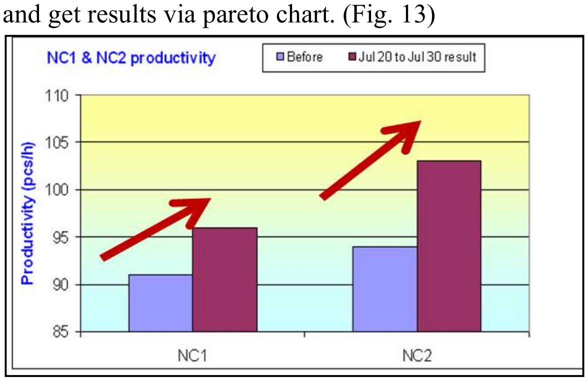


Figure 12: Daily report chart

6. Results

After doing some action to improve in the process and get results via pareto chart. (Fig. 13)



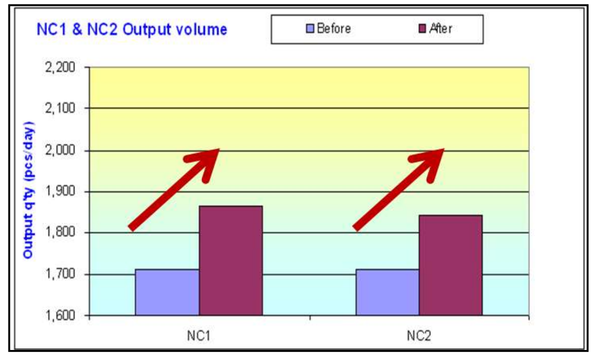


Figure 13: Output of NCL

Above charts show: Productivity was increased: NC1: 5.5% from 91 pcs/h to 96 pcs/h. NC2: 9.6% from 94 pcs/h to 103 pcs/h. Output quantity per day was increased: NC1: 8.8% from 1713 pcs/day to 1864 pcs/day. NC2: 7.6% from 1712 pcs/day to 1843 pcs/day.

7. Conclusion

The NC Latch process, under the study, faces many issues related to productivity loss. By DMAIC analysis the root causes were identified. The study identifies looking at the jig and stopping time as the major cause for loss of productivity. The causes and the recommended suggestions for both causes are discussed below.

Re-Layout and Re-Confirm operations were studied and it was observed that the pay structure and benefits were the factors that created dissatisfaction among the employees in the company. A productivity-based incentive scheme was recommended to the company that may help them to reduce stopping time for NC Latch machines and thus improve productivity.

8. Acknowledgment

I would like to thank all the employees and management of the organization under study for their immense contribution towards this research.

Reference

P. A. Vargas et al (2021), Towards industry improvement in manufacturing with DMAIC, *ICCIS 2020, AISC 1273*, pp. 341 – 352, 2021. Doi.org/10.1007/978-3-030-59194-6_28

L. Girmanova et al (2017), Application of six sigma using DMAIC methodology in the process of product quality control in metallurgical operation, *Acta Technologica Agriculturae 4*, 2017, pp.104-109, doi:10.1515/ata-2017-0020

A. N. S. NEJIB et al (2019), A new lean six sigma hybrid method based on the combination of PDCA and the DMAIC to improve process performance to clothing SME. *Industria textile*, 2019, vol.70 No. 5, doi: 10.35530/IT.070.05.1595

N. A. M. Salleh et al (2018), Cycle time improvement by lean six sigma for the cutting tool production line, *Journal of Mechanical Engineering,* Vol SI 5(I),240-264, 2018, ISSN: 1823-5514

Sunadi et al (2020), Analysis of aluminium alloy wheels product quality improvement through DMAIC method in casting process: a case study of the wheel manufacturing industry in Indonesia, *Journal Eudopeen des System Automatis*, Vol.54, No.1, February 2021. pp.55-62, doi:10.18280/jesa.540107

B. Phruksaphanrat et al., (2019), Six sigma DMAIC for machine efficiency improvement in a carpet factory, Songklanakarin J.Sci. Technol. 41(4),887-898, Jul – Aug.2019,

A. Shahin et al., (2010), Service Poka Yoke, *International Journal of Marketing Studies*, Vol.2, No.2; November 2010, ISSN: 1918-719X

H. Nishimura, (2015), Quality control andQualityimprovement, ISBN: 978-4-534-05320-6

Harukato, The Fundamentals of Production Managament, JMAM, 2003, ISBN: 4-8207-4069-5.

Ritsushi Tsukuda, (2012), A Practical guide of Toyota Production System, JMAM, ISBN: 978-4-8207-4788-8.

APPLICATION OF DATA ENVELOPMENT ANALYSIS IN LOGISTICS OUTSOURCING: THE CASE IN VIETNAM

Ngoc-Ai-Thy Nguyen[1,*], Thanh-Tuan Dang[1], Le-Thanh-Hieu Dang[2] and Chia-Nan Wang[1]

[1] Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

[2] Department of Industrial Management, Ho Chi Minh City University of Technology, Ho Chi Minh 72500, Vietnam

**Corresponding author: thy.logistics@gmail.com (N.-A.-T.N.)*

Abstract. The logistics cost accounts for about 25% of the average gross domestic product in Vietnam, while transport accounts for about 50-60% which is much higher than the rest of the world. In this context, outsourcing logistics to third-party logistics (3PL) providers is a critical component for today's businesses to lead supply chains toward sustainability. A complete transformation in business operations and logistical procedures caused by the Covid-19 pandemic indicates that processes have become more dynamic than ever before. It is imperative to implement performance evaluation of 3PLs to gain insights into their competitiveness and potential by observing efficiency indexes. This paper studies the top 25 logistics companies listed on the stock market in Vietnam and utilizes the data envelopment analysis (DEA) method, which are CCR-I, BBC-I, SBM-I-C, and EBM-I-C, for evaluating and selecting the most efficient and compatible 3PLs. In the DEA models, based on the importance of the financial indexes in the logistics industry, this paper considers four input variables, namely total assets, owner's equity, liabilities, and operating cost, and two output variables, namely sale revenue and net profit. The results show that Gemadept, Transimex, Vinalink, Vietnam Container, Saigon Logistics, Hai Phong Transport, and Green Investment achieved perfect efficiency scores in all models during the research period. This paper serves as a significant guideline for decision-makers, policymakers, and customers in their outsourcing strategy toward sustainable development.

Keywords: Third-party logistics; data envelopment analysis; efficiency score; decision making; sustainability.

1. Introduction

The world is currently fighting with global infectious pandemic known as Covid-19 which has started since December 2019 (World Health Organization, July 2020). Not only is the global health care system put in an emergency, but the global economy is also predicted to be suffered. To stop the spreading of the coronavirus, the limitation on human interactions is reported to be a feasible measure that is commonly applied across countries and regions (Hellewell et al., 2020). The main activities in this limitation can be listed as social distancing, lockdown, border closure, home quarantine, and school closures with different levels of compliance. Despite the effectiveness, the impact of these measures together with the current burden of the healthcare system has increased vividly in both micro and macroeconomics. Fernandes et al. (Fernandes, 2020) made a review and forecasted on the effect of this pandemic. Due to the uncertainty of lockdown duration and the recovery, the study predicts that the GDP growth can be decreased from 2.8% in mild scenarios and up to 15% for the critical situation. The study also discussed the industries at risk such as tourism, supply chain. As a result, each additional month of crisis is estimated to cost 2.5-3% of global GDP. In the combat against the Covid-19 crisis and the threat of another global economic recession, the logistics industry is one of the risky industries that need to be put under high consideration.

Manufacturers, distribution firms, and those with supply chains have all been shown to benefit from outsourcing to third-party logistics (3PL) providers. The growth of the 3PL service market has been accelerated due to the resonance of the e-commerce boom and increased reverse logistics operations. Choosing a 3PL is a huge decision, thus a significant problem for any firm (Pamucar et al., 2019). Therefore, this article aims to provide a comprehensive and coherent model for assessing and ranking 3PL providers using data envelopment analysis (DEA). The DEA models including CCR-I, BBC-I, SBM-I-C, and EBM-I-C are used to scale down the list of potential 3PLs. Eventually, the author considered 25 providers of Vietnam into the evaluation and selection process by using the prescribed approach. The developed integrated model can assist management stakeholders in better understand the entire process of 3PL assessment and selection from the standpoint of environmental sustainability. As a result, the evolved approach offers a complete, accurate, and robust decision-making method for the 3PL evaluation and selection.

Developed by Charnes et al. (1978), DEA is a powerful data processing and analysis tool for measuring the efficiency of DMUs (decision-making units), which are defined by the output-to-input ratio of a system. The aim of the DEA approach is to select a set of efficient and inefficient units. In the past few decades, DEA was applied in many industries such as banking, education, aviation, energy, to name a few (Wang et al., 2021; Bougnol et al., 2006; Fethi et al., 2000; Wang et al., 2021). CCR (Charnes–Cooper–Rhodes) and BCC (Banker–Charnes–Cooper) are the first models of DEA. CCR is the most basic model which measures technical efficiency of DMUs based on weights applied to inputs and outputs. BBC model is based on CCR with adjustments to address the constraints in the CCR model. Regarding the slack-based measure (SBM) model, it is the second adjustment to the CCR model with consideration to the slacks in the input and output factors. Slacks represent the potential improvement in the input and output variables for the inefficient units in the data set when compared with their peer efficient ultimate benchmark target. Proposed by Tone and Tsutsui (2010), the epsilon-based measure (EBM) is an epsilon-based DEA model that examines the DMUs' efficiency and inefficiency in a unified framework that combines radial and non-radial models. EBM was invented as a solution for the shortcomings in CCR, BCC and SBM models, which combines both radial and non-radial features. In the EBM model, a scalar measures epsilon that represents the diversity or the scattering of the observed dataset.

Thus, the authors aim to fill the gap in the existing literature by solving a case study of evaluating and selecting 25 potential 3PL providers in Vietnam. The main contributions of this paper is as follows: (1) a novel combinational DEA models is devoted to the emerging research topic of sustainability in the 3PL market; (2) a case study of assessing 25 potential 3PL providers in Vietnam is used to test the quality of the model; (3) the results reflect that the gap of applying the proposed method in the field of the 3PL industry is successfully addressed, and the integrated framework can be a powerful decision support tool for the evaluation and selection problem in any area.

2. Methodology

2.1. Charnes-Cooper-Rhodes model (CCR)

This section presents a brief mathematical model of data envelopment analysis (DEA). The list of symbols and notations below is used in the CCR, BCC and SBM model.

n: number of decision-making units (DMUs)
DMU$_i$: the *i-th* DMU, $i = 1, 2, \ldots, n$
DMU$_0$: the DMU target
$a_0 = (a_{01}, a_{02}, \ldots, a_{0p})$: input vector of DMU$_0$
$b_0 = (b_{01}, b_{02}, \ldots, b_{0q})$: output vector of DMU$_0$
$a_i = (a_{i1}, a_{i2}, \ldots, a_{ip})$: input vector of DMU$_i$, $i = 1, 2, \ldots, n$
$b_i = (b_{i1}, b_{i2}, \ldots, b_{iq})$: output vector of DMU$_i$, $i = 1, 2, \ldots, n$
$u \in R^{p\times 1}$: weight-input vector
$v \in R^{q\times 1}$: weight-output vector

CCR is the initial DEA models by Charnes et al. (1978). The procedure of CCR input-oriented (CCR-I) is defined in model (1).

$$\begin{aligned} &\underset{u,v}{Max}\, \xi = \frac{v^T b_0}{u^T a_0} \\ &\text{such that} \\ &v^T b_e \le u^T a_e, e = 1, 2, \ldots, n \\ &u \ge 0, v \ge 0 \end{aligned} \tag{1}$$

2.2. Banker-Charnes-Cooper model (BCC)

Following CCR model, the procedure of BBC input-oriented (BBC-I) is introduced as a linear model (2), as follows, Wen (2015).

$$\begin{aligned} &\underset{u,v}{Max}\, \xi = \frac{v^T b_0 - v_0}{u^T a_0} \\ &\text{such that} \\ &\frac{v^T b_e - v_0}{u^T a_e} \le 1, e = 1, 2, \ldots, n \\ &u \ge 0, v \ge 0 \end{aligned} \tag{2}$$

2.3. Slacks-Based Measure model (SBM)

The SBM input-oriented (SBM-I-C) under constant returns-to-scale assumption. The linear model is presented, as can be seen in model (3) as below, Farrell (1957).

$$\begin{aligned} &\omega_{In}^* = \underset{\alpha, s^-, s^+}{Min}\, 1 - \frac{1}{p}\sum_{i=1}^{p} \frac{s_i^-}{a_{i0}} \\ &\text{such that} \\ &\sum_{e=1}^{n} a_{ie}\alpha_e = a_{i0} - s_i^-, i = 1, 2, \ldots, p \\ &\sum_{e=1}^{n} b_{re}\alpha_e = b_{r0} + s_r^+, r = 1, 2, \ldots, q \\ &\alpha_e \ge 0, e = 1, 2, \ldots, n \\ &s_i^- \ge 0, i = 1,2, \ldots, p \end{aligned} \tag{3}$$

$$s_r^+ \geq 0, r = 1,2, \ldots, q$$

where ω_{In}^* denotes SBM-I-C efficiency.

2.4. Epsilon-Based Measure model (EBM)

Tone and Tsutsui (2010) proposed epsilon-based measure (EBM) as a solution for the shortcoming of CCR and SBM model. The EBM model considers n DMUs ($j = 1, 2, \ldots, n$) including m inputs ($i = 1, 2, \ldots, m$) and s outputs ($r = 1, 2, \ldots, s$). $X = \{x_{ij}\} \in R^{m \times n}$ and $Y = \{y_{rj}\} \in R^{s \times n}$ represent input and output matrices, respectively. The matrix X and Y are non-negative. The input-oriented model with a constant return to scale (EBM-I-C) is shown, as can be seen in model (4).

$$\delta^* = \underset{\theta,\lambda,s^-}{Min}\, \theta - \varepsilon_x \sum_{i=1}^{m} \frac{w_i^- s_i^-}{x_{io}}$$

such that

$$\sum_{j=1}^{n} x_{ij}\lambda_j = \theta x_{io} - s_i^-, i = 1, .., \mathrm{m}$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{ro}, r = 1, .., \mathrm{s} \qquad (4)$$

$$\lambda_j \geq 0, j = 1, 2, \ldots, n$$

$$s_i^- \geq 0, i = 1,2, \ldots, \mathrm{m}$$

where λ_j denotes the intensive vector of DMU, the subscript "o" denotes the DMU is under evaluation, s_i^- and w_i^- denote the amount of slack and weight in the i^{th} input, a parameter ε_x which depends on the degree of dispersion of inputs, and θ denotes the radial properties.

3. Case study

The logistics industry in Vietnam is currently facing several challenges including poor infrastructure, shortage of qualified labor force, unfavorable institutions and policies, complicated administrative procedures, and high logistics costs. Third-party logistics (3PL) are a well-liked method to outsource logistic functions, especially for small and medium-sized businesses. The strategic selection of 3PL can help the companies to reduce costs, boost sales and build brand identity with other competitors (Vietnam logistics industry, July 2021). In this paper, a case study of the top 25 logistics providers (DMUs) in Vietnam is chosen to demonstrate the effectiveness of the proposed model. The dataset of 25 DMUs is collected from the annual report of logistics companies in Vietnam's stock market (Vietnamese stock market, June 2021). The list of Vietnamese logistics companies and their profit in 2019 is as in Table 1.

As the proposed methodologies, DEA models including CCR-I, BBC-I, SBM-I-C, EBM-I-C are used to choose the candidate potential logistics providers from 25 DMUs. In the DEA model, the selection of inputs and outputs will directly affect the accuracy of the results. Based on the importance of the financial indexes in the logistics industry in literature, this paper considers four inputs (total assets, owner's equity, liabilities, operating cost) and two outputs (sale revenue, net profit), as can be seen in Figure 1. Following that, the descriptive statistics of inputs and outputs dataset (i.e., maximum, minimum, average, standard deviation), which is collected in Vietnam's stock market in 2019 (the unit is calculated in thousands of US dollars, are shown in Table 2.

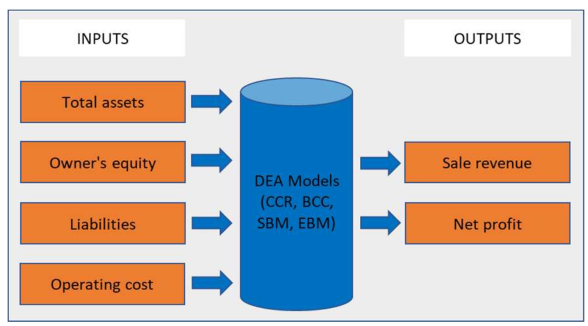


Figure 1. Inputs and outputs of DEA models.

Table 1. List of logistics companies and their profit in 2019 (unit: thousand VND).

DMUs	Company	Symbol	Code	Profit
3PL-01	Gemadept Corporation	Gemadept	GMD	613,569
3PL-02	Transimex Corporation	Transimex	TMS	225,324
3PL-03	Ha Tien Transport JSC	Ha Tien Transport	HTV	29,377
3PL-04	South Logistics JSC	South Logistics	STG	122,918
3PL-05	Duyen Hai Multi Modal Transport JSC	Duyen Hai Transport	TCO	16,251
3PL-06	Vinalink Logistics JSC	Vinalink	VNL	18,788
3PL-07	Vietnam Container Shipping JSC	Vietnam Container	VSC	285,795

3PL-08	Viet Nam Tanker JCS	Vietnam Tanker	VTO	99,285
3PL-09	ASG Corporation	ASG Logistics	ASG	156,368
3PL-10	Hai An Transport and Stevedoring JSC	Hai An Transport	HAH	132,739
3PL-11	Superdong Fast Ferry Kieng Giang JSC	Kien Giang Logistics	SKG	101,100
3PL-12	Saigon Ground Services JSC	Saigon Logistics	SGN	378,524
3PL-13	SCSC Cargo Service Corporation	SCSC Logistics	SCS	68,122
3PL-14	MHC JSC	MHC Logistics	MHC	9,045
3PL-15	Ben Thanh Service JSC	Ben Thanh Logistics	BSC	554
3PL-16	Hai Phong Cement Transport and Trading JSC	Hai Phong Transport	HCT	1,947
3PL-17	Hai Minh Corporation	Hai Minh Logistics	HMH	11,896
3PL-18	Vinafreight JSC	Vinafreight	VNF	19,795
3PL-19	The Van Cargoes and Foreign Trade Logistics JSC	Van Cargo	VNT	4,106
3PL-20	West Coach Station JSC	West Coach	WCS	11,953
3PL-21	Green Development and Investment Service JSC	Green Investment	GIC	42,667
3PL-22	Hoang Ha JSC	Hoang Ha Logistics	HHG	514
3PL-23	Portserco Logistics JSC	Portserco	PRC	1,435
3PL-24	Lien Ninh Transport and Service JSC	Lien Ninh Transport	BLN	2,193
3PL-25	Inland Waterways Management and Maintenance JSC	Inland Logistics	DT4	1,550

Table 2. Descriptive statistics of inputs and outputs.

	Minimum	**Maximum**	**Mean**	**Std. Deviation**
Total assets	1,898.50	43,9709.19	54,406.21	88,775.75
Owner's equity	573.58	285,346.82	37,077.41	58,114.35
Liabilities	186.05	154,362.37	17,170.65	32,266.78
Operating cost	70.43	21,819.68	3,908.52	5,772.31
Sale revenue	181.88	114,834.41	34,623.30	34,369.48
Net profit	22.33	26,659.53	4,094.40	6,372.18

4. Results analysis

Before applying the DEA models, the Pearson correlation coefficients must be checked to satisfies the isotonicity and homogeneity conditions of input and output factors. Table 3 shows the results of Pearson coefficient of inputs and outputs, from the results, the correlation value ranges from 0.590 to 1, which shows a positive linear relationship. Hence, the dataset can be used for the DEA models. Note: ** denotes correlation is significant at the 0.01 level (2-tailed).

Table 4 shows the candidate potential 3PL from after running four DEA models. From this result, there are seven potential DMUs, which are Gemadept (3PL-01), Transimex (3PL-02), Vinalink (3PL-06), Vietnam Container (3PL-07), Saigon Logistics (3PL-12), Hai Phong Transport (3PL-16), Green Investment (3PL-21). These potential DMUs achieve the DEA efficiency score of 1 or 100% for all four models including CCR-I, BBC-I, SBM-I-C, and EBM-I-C (i.e., the optimal DMUs). Otherwise, they are considered as DEA inefficiency. Figure 2 shown the results comparison among DEA models, the visualized chart presented that the efficiency score values are ascended as SBM-I-C > EBM-I-C > CCR-I > BCC-I.

Table 3. Pearson coefficient of inputs and outputs.

		TA	OE	LI	OC	SR	NP
Total assets (TA)	Pearson coefficient	1	0.990**	0.965**	0.678**	0.770**	0.881**
	P-value		0.000	0.000	0.000	0.000	0.000
	Sample size	25	25	25	25	25	25
Owner's equity (OE)	Pearson coefficient	0.990**	1	0.919**	0.708**	0.759**	0.901**
	P-value	0.000		0.000	0.000	0.000	0.000
	Sample size	25	25	25	25	25	25
Liabilities (LI)	Pearson coefficient	0.965**	0.919**	1	0.590**	0.746**	0.795**
	P-value	0.000	0.000		0.002	0.000	0.000
	Sample size	25	25	25	25	25	25
Operating cost (OC)	Pearson coefficient	0.678**	0.708**	0.590**	1	0.597**	0.609**
	P-value	0.000	0.000	0.002		0.002	0.001
	Sample size	25	25	25	25	25	25
Sale revenue (SR)	Pearson coefficient	0.770**	0.759**	0.746**	0.597**	1	0.796**
	P-value	0.000	0.000	0.000	0.002		0.000
	Sample size	25	25	25	25	25	25
Net profit (NP)	Pearson coefficient	0.881**	0.901**	0.795**	0.609**	0.796**	1
	P-value	0.000	0.000	0.000	0.001	0.000	
	Sample size	25	25	25	25	25	25

Table 4. Candidate potential 3pls from DEA models.

No.	DMUs	Symbol	CCR-I	BCC-I	SBM-I-C	EBM-I-C
1	3PL-01	Gemadept	1	1	1	1
2	3PL-02	Transimex	1	1	1	1
3	3PL-03	Ha Tien Transport	0.53805	0.58660	0.42496	0.51658
4	3PL-04	South Logistics	0.39367	0.92567	0.35704	0.39094
5	3PL-05	Duyen Hai Transport	0.68594	0.70078	0.59022	0.68487
6	3PL-06	Vinalink	1	1	1	1
7	3PL-07	Vietnam Container	1	1	1	1
8	3PL-08	Vietnam Tanker	0.57517	0.98440	0.42028	0.51963
9	3PL-09	ASG Logistics	0.47798	0.47920	0.43595	0.46748
10	3PL-10	Hai An Transport	0.65784	0.71271	0.44272	0.64153
11	3PL-11	Kien Giang Logistics	0.81947	1	0.50706	0.76352
12	3PL-12	Saigon Logistics	1	1	1	1
13	3PL-13	SCSC Logistics	0.60740	1	0.45332	0.56492
14	3PL-14	MHC Logistics	0.39364	0.57313	0.14755	0.31117
15	3PL-15	Ben Thanh Logistics	0.11948	1	0.09068	0.11497
16	3PL-16	Hai Phong Transport	1	1	1	1
17	3PL-17	Hai Minh Logistics	0.78664	0.90565	0.55693	0.72221
18	3PL-18	Vinafreight	0.78252	1	0.63822	0.74954
19	3PL-19	Van Cargo	0.91732	1	0.66386	0.84657
20	3PL-20	West Coach	0.28526	0.29335	0.24404	0.27393
21	3PL-21	Green Investment	1	1	1	1

22	3PL-22	Hoang Ha Logistics	0.26264	0.34832	0.19380	0.23710
23	3PL-23	Portserco	0.74086	1	0.58021	0.71495
24	3PL-24	Lien Ninh Transport	0.62335	0.67417	0.45494	0.58861
25	3PL-25	Inland Logistics	0.39650	1	0.25712	0.378822

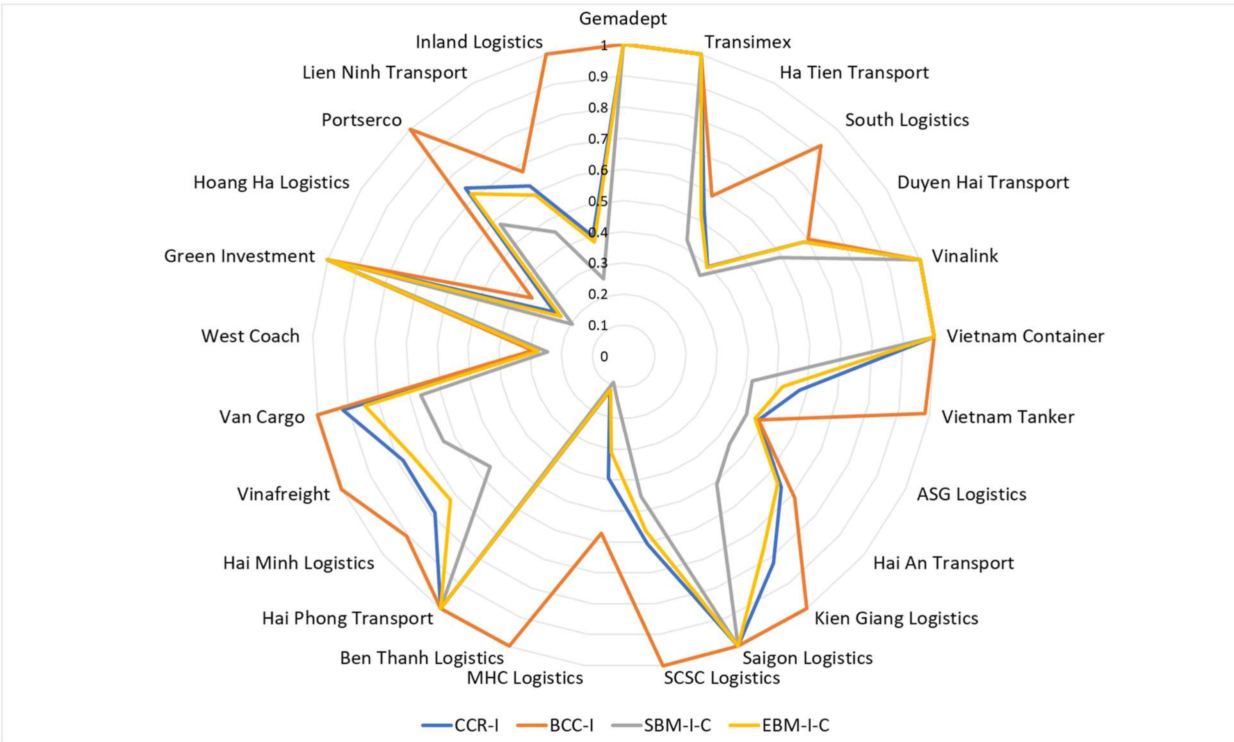


Figure 2. The results comparison among DEA models.

5. Conclusions

Efficient and momentous decisions on outsourcing are vital to the performance of businesses in today's competitive market. A complete transformation in business operations and logistical procedures caused by the Covid-19 pandemic indicates that processes have become more dynamic than before, so that companies are being placed in a position to choose between the most critical tasks. The third-party logistics providers (3PL) evaluation and selection can be analyzed as a multiple-criteria decision-making problem (MCDM). This paper applied DEA models (CCR-I, BBC-I, SBM-I-C, and EBM-I-C) for solving the logistics outsourcing problems to the Vietnamese third party in this era. The results show that Gemadept, Transimex, Vinalink, Vietnam Container, Saigon Logistics, Hai Phong Transport, and Green Investment achieved perfect efficiency scores in all models during the research period. Furthermore, the managerial implications of this paper are to support the decision-makers to aware of their organization's position in the logistics industry, to apprehend overall efficiency and productivity. Future studies should consider new assessment factors that can impact the process of evaluation and selection of logistics providers to enhance the robust results.

6. References

Bougnol, M. L., & Dulá, J. H. (2006). Validating DEA as a ranking tool: An application of DEA to assess performance in higher education. *Annals of Operations Research*, **145**, 339-365.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European journal of operational research*, **2**, 429-444.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, **120**, 253-281.

Fernandes, N. Economic effects of coronavirus outbreak (COVID-19) on the world economy 2020. Available at SSRN 3557504, 2020.

Fethi, M. D., Jackson, P. M., & Weyman-Jones, T. G. (2000). Measuring the efficiency of European airlines: an application of DEA and Tobit Analysis.

Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W., ... & Eggo, R. M. (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, **8**, e488-e496.

Pamucar, D., Chatterjee, K., & Zavadskas, E. K. (2019). Assessment of third-party logistics provider using multi-criteria decision-making approach based on interval rough numbers. *Computers & Industrial Engineering*, **127**, 383-407.

Tone, K., & Tsutsui, M. (2010). An epsilon-based measure of efficiency in DEA–a third pole of technical efficiency. *European Journal of Operational Research*, **207**, 1554-1563.

Vietnam logistics industry: risks and challenges. Available online: https://vietnamcredit.com.vn/news/vietnam-logistics-industry-risks-and-challenges_13632 (accessed on 06 July 2021).

Vietnamese stock market. Available online: https://vietstock.vn/ (accessed on 03 June 2021).

Wang, C. N., Nguyen, N. A. T, Dang, T. T., & Bayer, J. (2021). A Two-Stage Multiple Criteria Decision Making for Site Selection of Solar Photovoltaic (PV) Power Plant: A Case Study in Taiwan. *IEEE Access*, **9**, 75509-75525.

Wang, C. N., Nguyen, N. A. T., Dang, T. T., & Trinh, T. T. Q. (2021). A Decision Support Model for Measuring Technological Progress and Productivity Growth: The Case of Commercial Banks in Vietnam. *Axioms*, **10**, 131.

Wen, M. Uncertain Data Envelopment Analysis. In Uncertainty and Operations Research, Springer: Heidelberg, Berlin, 2015.

World Health Organization. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 06 July 2020).

EXAMINING THE BARRIERS ASSOCIATED WITH AGILE MANUFACTURING IMPLEMENTATION: CASE STUDIES IN VIETNAM

Ton Nguyen Trong Hien[1]

[1] *School of Business Administration, Van Lang University, hien.tnt@vlu.edu.vn*

Abstract. The business climate has become more complex due to an increasingly dynamic and competitive environment. This new competitive landscape requires manufacturers to have taken into account various complex factors globalization causes. One of these umbrella groupings is Agile manufacturing (AM), a strategy that incorporates within itself flexibility. A few previous studies have tried to investigate possible barriers to adopting Agile production, however, the barriers have not been explored in Vietnam. To obtain a better view of problematic areas in Agile transformation, the main objective of this paper is to explore barriers associated with AM in Vietnamese manufacturing context. The study aims to help the decision-maker to anticipate potential obstacles and take proper measures to deal with them during Agile implementation.

Keywords: Agile manufacturing, Barriers, Vietnam.

1. Introduction

Historically, manufacturers have invested heavily in resources for improving efficiency and reducing the cost of production (Hagel III et al., 2015; Mohanty & Prakash, 2014), to support growth and success within an ever-changing and increasingly competitive global economy. To remain successful, organizations must be able to adapt quickly to unpredictable changes, thus applying pressure for implementing continuous improvement initiatives that optimize cost, quality, and efficiency. However, these requirements have subsequently introduced many challenges to overcome, such as the unpredictability of customer demand. This instability in demand may threaten the responsiveness of formal material and production planning systems and thus, has subsequently resulted in many manufacturing organizations seeking an effective method towards delivering improved manufacturing operations to fend off competition. One method for overcoming such challenges is Agile Manufacturing, which was introduced as a new paradigm in the early 1990s (Gunasekaran et al., 2018).

Several researchers believe that AM can enable organizations to respond quickly to sudden changes in market demand. Maskell (2001) suggests that the key elements of the paradigm are customer prosperity, people and information, co-operation within and between companies, and adapting for change. Whereas Thilak et al. (2015) view the critical characteristics of AM being in the manufacturing system's re-configurability and flexibility to deliver a diverse range of products. Accordingly, in light of unprecedented market instability compounded by unpredictable customer requirements, Aravind Raj et al. (2013) and Gunasekaran et al. (2018) identify the following "agility enablers and practices" as a means to deliver a competitive advantage through AM principles; agile supply chain, total employee empowerment, transparent or mass customization, management responsibility, intelligent automation and information technology integration. While recent years have shown increasing use of AM, these efforts however have not been successful, i.e., attempts to implement Agile often fail (Thilak et al., 2015).

Evaluating AM barriers has been studied in a few places; however, research results vary from country to country, and practical research in each country may not be the same. Also, the production process is slow in developing countries. The AM paradigm is relatively new (compared to Lean), and it has not been investigated in Vietnam, therefore it is necessary to research and assess the barriers to implementing AM in Vietnamese manufacturing enterprises.

2. Research Methodology

This study employed a combined descriptive and inferential research design methodology, which was conducted in three stages (Figure 1). As a first step, we attempted to recognize barriers in the implementation of AM; these were identified via a literature review of previous studies to avoid bias. The second step involved collecting primary data through interviewing a group of experts to assess the applicability of the barriers (identified in stage 1) in a Vietnamese manufacturing context. The third and final step involved the analysis of data gained from the interviewing process to determine the significance of each barrier, in the context of Vietnamese manufacturing organizations which are implementing AM.

Figure 1: Steps of the research process

2.1. Literature review

To ensure the reliability and credibility of the literature used, we relied upon searching both Scopus and Web of Science (WoS) databases for peer-reviewed articles only. After entering the search key term "(TITLE (barrier) AND TITLE (agile AND production) OR TITLE (agile AND manufacturing))" on September 10, 2020, the results were as follows:

A total of five articles were returned via the Web of Science database (2 articles from 2017, 2 from 2019, and 1 article from 2020), while the Scopus database returned ten results (3 articles from 2020, 2 from 2019, 1 from 2018, 2 from 2017, 1 from 2016, and 1 article from 2007). Screening all fifteen articles against their titles, authors, and published dates highlighted five duplicate articles, thus resulting in a total of ten peer-reviewed articles.

The analysis showed that most of the authors come from India (with some US authors also joining the Indian study group). To ensure the selection of high-quality journals contained within the databases, we proceeded to further refine the screening process by considering the citation index and impact factor (IF) of each article. Although there is a consensus that impact factors should not be discarded (Bornmann & Pudovkin, 2017), many suggest that citation count and journal impact factors are not suitable for the evaluation of research (Seglen, 1998).

Many elements affect the impact factor, not only the quality of the article but also its readability and accessibility to readers, etc. (Tahamtan et al., 2016). The same goes for citation indexes, which are increasingly used as performance indicators in research policy. Normally, citations are assumed to reflect the impact of the study or the quality of the article. However, there is no evidence to suggest that the citations reflect aspects of research quality (Aksnes et al., 2019). Therefore, as there is no clear evidence in the use of citation index or journal impact factor as a measure of reliable literature, the selection of articles included within this study are based on the AJG 2018 standard, as recommended by the Chartered Association of business school (UK) to encompass the highest quality production-oriented journals. The results of the selection are shown in Table 1, below:

Table 1: Summary of search results

No.	**DOI**	**Source**	**AJG 2018**
1	10.1108/TQM-04-2020-0073	TQM Journal	
2	10.1007/s00170-020-05486-5	International Journal of Advanced Manufacturing Technology	
3	10.1007/978-981-15-4565-8_12	Lecture Notes in Mechanical Engineering	
4	10.1108/BIJ-09-2017-0245	Benchmarking	
5	10.1007/978-981-13-6412-9_62	Lecture Notes in Mechanical Engineering	
6	http://www.iaeme.com/ijmet/issues.asp? JType=IJMET&VType=9&IType=11	International Journal of Mechanical Engineering and Technology	
7	10.1109/AMIAMS.2017.8069180	AMIAMS 2017 – Proceedings	
8	10.1108/BIJ-02-2016-0024	Benchmarking	
9	10.1504/IJASM.2016.081558	International Journal of Agile Systems and Management	x
10	10.1504/IJASM.2007.015679	International Journal of Agile Systems and Management	x

Of the ten focal articles, only two conformed to the AJG 2018 standard. Sindhwani and Malhotra (2016) highlighted three barriers that manufacturing enterprises in India are facing when implementing AM, while Hasan et al. (2007) analyzed the relationships between eleven barriers. After aggregating the idea of duplication, we have gathered twelve barriers that manufacturing businesses face when applying AM:

Table 2. Barriers to Agile implementation

No.	Barriers	Description
1	Poor design manufacturing process	Delay in the manufacturing process
2	Defective parts production	Every time that we must rework product according to the required specification
3	Lack of methodology to enhance agility	There is no Agile culture or guide to the implementation of Agile t hinking
4	Lack of top management support	The absence of top management support for continuous improvement
5	Resistant to change	Converting to AM is great but it can be difficult to implement because people are often resistant to change
6	Inappropriate measurement	We cannot measure how quickly and easily in response to changing customer demand, does not provide a sufficient metric for agility
7	Lack of customer feedback integration	No business processes or technology satisfy customer's needs immediately because integrating customer feedback can get to the root causes of business challenges faster
8	Insufficient training	Inadequately trained employees, unskilled workers, or employees who do not understand how to do their jobs
9	No investment appraisal techniques	No investment appraisal technique is used for advanced manufacturing technology
10	Poor partnership	Do not have supply chain and network relationships
11	Unavailability of appropriate technology	Outdated or inappropriate technology
12	Lack of flexibility	Flexibility incorporation occurs at strategic and operational levels

2.2. Primary data collection

This study adopts a questionnaire interview methodology to explore the significant barriers associated with AM implementation in Vietnam. The criteria set out for the interview are described as follows:

- *Timespan*: 30/9/2020 – 10/01/2021
- *Scope*: The interview was conducted with six production plants operate the following fields in Vietnam: textile, knife, and fork production, food and beverage, electrical switchboard manufacturing.
- *Means of the interview*: face-to-face interview, online interview via Google form, and phone interview.
- *Interview audiences*: they are diverse, those are: production staff, production director, project management staff, QA staff, QA manager, logistics staff, supply chain manager, operations manager, R&D staff, R&D manager. Most of them are in high positions; this means their current state concerns the implementation of Agile.
- *Sample size*: 70 questionnaires were sent.
- *Interview process*: A total of 12 possible barriers to the implementation of AM have been investigated in this study. Besides, these challenges are assessed in this study by scoring items on a Likert-type scale with responses ranging from 1 "strongly disagree" to 5 "strongly agree" according to the degree of contribution.

Because the concept of AM is relatively new, the conductor of this study ensured to explain the definition of AM and clarified the interview objectives to all participants, prior to pretesting and piloting the questionnaire. Moreover, all participants were assured of the confidentiality surrounding their responses.

2.3. Analysis

Out of the 70 questionnaires distributed among the target respondents, 65 questionnaires were returned. The reliability of the collected data was calculated using Cronbach's Alpha method to provide a measure of the internal consistency of a test. Cronbach's α coefficient is a statistical test of the degree of rigor that the question items in the scale are correlated, where: $\alpha = N\rho \,/\, [1 + \rho\,(N-1)]$. This was carried out by employing the Statistical Package for Social Science (SPSS version 25) software to compute Cronbach Alpha values. Thus, all data collected via questionnaires were subjected to Cronbach's Alpha analyses (per Table 3) to estimate the internal consistency and reliability of each response, prior to data interpretation.

Test results for the 12 AM barrier variables evaluated in this study resulted in a Cronbach Alpha value of 0.878, thus indicating that this would be indicative of a very good internal consistency of the test. In other words, the figure demonstrates responses provided by the respondents, it is said to have very good reliability of 87.8%.

The Relative Importance Index (RII) was employed as it best fits the purpose of this study to rank the criteria according to their relative importance (Khatib & El-Shafie, 2020). Mathematically, the RII is obtained as follows:

$$\mathbf{RII} = \frac{\sum W}{A \times N}$$

where W, represents the summation of Likert allocation. A is the highest weight (i.e., 5 in this case) and N represents the total number of participants (65 for this study).

Table 3. Performance of Cronbach's alpha (DeVellis, 2017)

No.	Cronbach's alpha, α	Internal consistency
1	α ≥ 0.9	Consider shortening the scale
2	0.90 > α ≥ 0.80	Very good
3	0.80 > α ≥ 0.70	Good
4	0.70 > α ≥ 0.65	Satisfactory
5	0.65 > α ≥ 0.60	Undesirable
6	α < 0.60	Unacceptable

To determine the ranking of different factors, the barriers have been ranked based on Relative Importance Index (RII) and Mean Value. The RII and Mean value rankings were classified into their level of effect in Table 4.

Table 4. Classification of RII introduced (Darus et al. 2015)

Scale	Contribution level	RII
1	Very low	0.0 ≤ RII ≤ 0.2
2	Low	0.2 < RII ≤ 0.4
3	Average	0.4 < RII ≤ 0.6
4	High	0.6 < RII ≤ 0.8
5	Very high	0.8 < RII ≤ 1.0

3. Research Result

Through an interview and analysis of the perceptions of respondents, Table 5 shows the level of contribution of the top three challenges faced when implementing Agile production is found to be very high. As stated in Table 5, the poor design manufacturing process (RII = 0.957) is the top barrier confronted that prevents the successful adoption of AM. Poor partnership and interactions between vendors are ranked second in the group with RII = 0.846, lack of flexibility (RII = 0.815) is ranked third, lack of customer feedback integration (RII = 0.782) is ranked fourth, absence of top management commitment (RII = 0.769) is ranked fifth, unavailability of appropriate technology (RII = 0.646) is ranked sixth, insufficient training or education (RII = 0.640) is ranked seventh, lack of methodology to enhance agility (RII = 0.631) is ranked eighth, an inappropriate measurement for agility (RII = 0.6250) is ranked ninth, defective parts production (RII = 0.554) is ranked tenth, no investment appraisal techniques (RII = 542) is ranked eleventh, while resistant to change (RII = 0.397) is ranked twelve.

It shows that most of the respondents are taking awareness of the situation that diminishes the productivity of the manufacturing systems. However, respondents argue that the barrier to an AM enterprise is not from production delay alone but rather the delay in the entire supply chain. Therefore, this issue raises concerns about flexibility in management, sharing information across all supply chain partners. Furthermore, there is a need to focus on strengthening relationships in the supply chain, because "partnership will make the leadership team stronger and more agile" (Taylor & Haneberg, 2010).

Table 5. Most Important Factors casing Agile manufacturing barriers

Categorization of Agile manufacturing barriers in Vietnam	RII	Mean value	RII & Mean value	Contribution level
Poor design of manufacturing process	0.957	4.785	1	**Very high**
Defectives parts production	0.554	2.769	10	**Average**
Lack of methodology to enhance agility	0.631	3.154	8	**High**
Lack of top management support	0.769	3.846	5	**High**
Resistant to change	0.397	1.985	12	**Low**
Inappropriate measurement	0.625	3.123	9	**High**
Lack of customer feedback integration	0.782	3.908	4	**High**
Insufficient training	0.640	3.200	7	**High**
No investment appraisal techniques	0.542	2.708	11	**Average**
Poor partnership	0.846	4.231	2	**Very high**
Unavailability of appropriate technology	0.646	3.231	6	**High**
Lack of flexibility	0.815	4.077	3	**Very high**

4. Conclusion

This research aims to examine the significant barriers that derail the implementation of AM. A total of 12 barriers has been defined and analyzed to successfully reach the research goal. Like a house that needs to be built from the ground up, AM requires a standard manufacturing process, strong partnership, and requires special attention to its implementation and removal of its barriers from strategic and operational levels. In the future, it will be necessary to explore these barriers for a specific production sector.

5. Reference

Aksnes, D., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*, 9(1), 215824401982957.

Aravind Raj, S., Sudheer, A., Vinodh, S., & Anand, G. (2013). A mathematical model to evaluate the role of agility enablers and criteria in a manufacturing environment. *International Journal of Production Research*, 51(19), 5971-5984.

Bornmann, L., & Pudovkin, A. (2017). The Journal Impact Factor Should Not Be Discarded. *Journal of Korean Medical Science*, 32(2), 180-194.

Darus, N., Haron, Z., Mohd Bakhori, S., Ming Han, L., Jahya, Z., & Abdul Hamid, M. (2015).

Construction Noise Annoyance among the Public Residents. *Jurnal Teknologi*, 74(4).

DeVellis, R. (2017). *Scale development* (4th ed.). Los Angeles [etc.]: SAGE.

Gunasekaran, A., Yusuf, Y., Adeleye, E., Papadopoulos, T., Kovvuri, D., & Geyi, D. (2018). Agile manufacturing: an evolutionary review of practices. *International Journal of Production Research*, 57(15-16), 5154-5174.

Hagel III, J., Brown, J. S., Kulasooriya, D., Giffi, C. and Chen, M. (2015), The future of Manufacturing-Making things in a changing world. *Future of the Business Landscape. Deloitte University Press*., Westlake, USA.

Hasan, M., Shankar, R., & Sarkis, J. (2007). A study of barriers to agile manufacturing. *International Journal of Agile Systems and Management*, 2(1), 1.

Khatib, B., Poh, Y., & El-Shafie, A. (2020). Delay Factors Management and Ranking for Reconstruction and Rehabilitation Projects Based on the Relative Importance Index (RII). *Sustainability*, 12(15), 6171.

Maskell, B. (2001). The age of agile manufacturing. *Supply Chain Management: An International Journal*, 6(1), 5-11.

Mohanty, R., & Prakash, A. (2014). Green supply chain management practices in India: a confirmatory empirical study. *Production & Manufacturing Research*, 2(1), 438-456.

Seglen, P. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthopaedica Scandinavica*, 69(3), 224-229.

Sindhwani, R., & Malhotra, V. (2016). Barrier's evaluation for agile manufacturing system with fuzzy performance importance index approach. *International Journal of Agile Systems and Management*, 9(4), 292.

Taylor, J., & Haneberg, L. (2010). *Connecting top managers* (p. 60). Upper Saddle River, N.J.: FT Press.

Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107(3), 1195-1225.

Thilak, V., Devadasan, S., & Sivaram, N. (2015). A Literature Review on the Progression of Agile Manufacturing Paradigm and Its Scope of Application in Pump Industry. *The Scientific World Journal*, 2015, 1-9.

TWO-STAGE STOCHASTIC PROGRAMMING FOR PRODUCTION PLANNING

Phan Nguyen Ky Phuc[1]

[1] *International Univeristy-Vietnam National University HoChiMinh City, Quarter 6, Linh Trung Ward, Thu Duc District, pnkphuc@hcmiu.edu.vn*

Abstract. Stochastic demand is an important factor that heavily affects production planning. Under its effects, company activities such as purchasing, manufacturing, selling, etc., are all considerably influenced and need to adapt quickly with its realization. In production planning, for several reasons, many decisions must be made in intial stage when the demand has not been realized. If these decsions are non-optimal, they will propagate their effects to the later stage which can lead to the severe loss for the company due to either serious over stocks or stock out. In order to find the optimal solutions for both the first stage and the second stage regarding to demand realization, this paper proposes a stochastic two-stage linear programming model for purchasing and production planning process for multi-supplier, multi-material and multi-product. The objective function is to maximize the expected profit earned back after two stages while the results include detailed plans on the purchasing and production regarding to each demand scenario)

Keywords: Mixed integer programming, two-stage stochastic programming, production planning, order allocation.

1. Introduction

Stochastic demand is an important factor that strongly drives the entire company operations. Due to the demand uncertainty, several activities such as purchasing, manufacturing, etc., must be carefully planned to avoid severe shortage or overstocks which are the main causes of company loss. Purchasing and manufacturing are activities which are directly influenced by demand. This study proposes a two-stage stochastic linear programming model for order allocation and production planning for multiple materials, multiple products. The model strives to answer the question of how much purchased quantity should be placed, a production planning table for each production stage providing basic information such as labor, operational hour, production line needed to produce.

Next section provides crucial literature review relating to the academic background of this research. Section 3 presents the proposed two-stage stochastic mixed-integer programming mathematical models while the conclusions are drawn in section 4.

2. Literature review

Stochastic programming is a well-established method in operations research and has been applied to many problems with uncertain parameters (Birge and Louveaux, 2011). The uncertain parameters are represented as random variables, where each possible combination of values constitutes one scenario. The model in this study follows the standard form of two-stage stochastic linear program with fixed recourse

$$\min_{x} c^T x + \mathcal{Q}(x)$$
$$Ax = b, x \geq 0$$

where $\mathcal{Q}(\mathrm{x}) \coloneqq \mathbb{E}[\mathrm{Q}(\mathrm{x}, \xi(\omega))]$ and $\mathcal{Q}(\mathrm{x}, \xi)$ is the optimal value of the second-stage problem.

$$\min_{\mathrm{x}} \mathrm{q}^{\mathrm{T}}\mathrm{y}$$

Since the problem is fixed recourse and its probability distribution, i.e, $\xi(\omega)$ has finite support; $\{\xi_1, \xi_2, \cdots, \xi_k\}$ with respective probabilities $p_k \in (0,1), k = 1,2, \cdots, K$. For such problems, the expected value function $\mathcal{Q}(\mathrm{x})$can be written as the finite sum

$$\mathcal{Q}(\mathrm{x}) = \sum_{k=1}^{K} p_k\, Q(x, \xi_k)$$

Then the problem can be combined into a single linear program as follows:

$$min_{x,y_1,y_2,\cdots,y_2} c^T x + \sum_{k=1}^{K} p_k q_k^T y_k$$
$$Ax = b, x \geq 0$$
$$W y_k = h_k - T x_k, y_k \geq 0, k = 1,2, \cdots, K$$

The main purpose of our model is to support decision makers in terms of order allocation and production planning to efficiently perform production and logistics. Based on this model several works relating to stochastic planning have been developed. Mahdavi et al. (2012) proposed a mixed integer linear programming to aggregate multi products in multi periods in considering various cost components. Altendorfer et al. (2016) discovered the effects of forecast error on optimizing the production plan and the total cost as well. Leung et al. (2006) proposed a two-stage stochastic programming using market demand variability to determine the optimal medium-term production plans, with the main objective function is to minimize the total expected production cost. Zanjani et al. (2009) used a multi-stage stochastic programming approach for production planning to minimize the total inventory and backorder costs for all products and materials. Leung & Ng (2007) optimized the production plan for perishable products under effect of stochastic market demand, production and inventory holding cost. Aouam & Uzsoy (2014) compared three different formulations for production planning problems such as chance constrained model, two-stage stochastic programming model and robust optimization model. Zanjani et al. (2011) studied about the production planning with uncertainty in the quality of raw materials in sawmill industry. Fang et al. (2017) considered a hybrid manufacturing and remanufacturing systems which reused remaufacturable parts in the end-of-use products. They investigated a production planning problem with the integration of resource capacity planning that is shared by both processes. Demands are assumed to be stochastic, and the problem was solved based on the Lagrangian relaxation technique and problem decomposition. Xinhui et al. (2011) proposed a stochastic production planning model for an international enclosure manufacturing company with seasonal demand and market growth uncertainty. The company forecasts demand and inform to suppliers to reduce risks for both parties. A two- stage stochastic production planning model is developed to minimize the total costs under all scenarios. Marteo et al. (2016) studied two-stage stochastic model for selecting grocery shop brand in order to ensure high quality products and minimal time from farm-to-table under seasonal contracts. The main aim is to minimize overall procurement costs and meet future demand. Their problem is common in fresh vegetable industry.

3. Model development

Annotation

m: index of materials $m = 1 \dots M$
p: index of products $p = 1 \dots P$
s: index of suppliers $s = 1 \dots S$
r: index of price ranges $r = 1 \dots R$
l: index of production lines $l = 1 \dots L$
e: index of scenarios $e = 1 \dots E$

Parameters

D_{pe}: demand of product type p under scenario e
$BigM$: very large number
α_{mp}: number of material m required for manufacturing one uint of product type p
β_m^M: savage price of one unit of material type m
β_p^P: savage price of one unit of product type p
γ_p: selling price of one unit of product type p
δ_p: Unit cost to manufacturing one unit of product type p
τ^1: Labour cost for one worker during stage 1
τ_e^2: Labour cost for one worker under scenario e during stage 2
θ_{msr}^1: the offer price of material type m from supplier s at price range r at stage 1
θ_{msre}^2: the offer price of material type m from supplier s at price range r under scenario e at stage 2
LB_{msr}^1: the lower bound for purchasing quantity of the material type m from supplier s at price range r in stage 1
LB_{msre}^2: the lower bound for purchasing quantity of the material type m from supplier s at price range r under scenario e in stage 2
UB_{msr}^1: the upper bound for purchasing quantity of the material type m from supplier s at price range r in stage 1
UB_{msre}^2: the upper bound for purchasing quantity of the material type m from supplier s at price range r under scenario e in stage 2
π_e: the probability associating to scenario e
Φ_{msr}^1: fixed cost for ordering material type m from supplier s at price range r at stage 1
Φ_{msre}^2: fixed cost for ordering material type m from supplier s at price range r under scenario e at stage 2
ω^1: fixed cost for using one line at stage 1
ω_e^2 : fixed cost for using one line under scenario e at stage 2
H_p^1 : number of available hours of one line for manufacturing a product type p at stage 1
H_{pe}^2 : number of available hours of one line for manufacturing a product type p under scenario e at stage 1
ϑ_p^1 : maximum number of production line for manufacturing product type p in stage 1
ϑ_{pe}^2 : maximum number of production line for manufacturing product type p under scenario e in stage 2
t_p: time required for manufacturing one product type p
C_{ms}^1: capacity of supplier s for material type m at stage 1
C_{mse}^2: capacity of supplier s for material type m under scenario e at stage 2
ψ: number of required workers for operating one production line

Decision variables

X_{msr}^1: number of material type m purchased from supplier s at price range r in stage 1
X_{msre}^2: number of material type m purchased from supplier s at price range r under scenario e in stage 2
Y_p^1 : number of product type p manufactured in stage 1
Y_{pe}^2: number of product type p manufactured under scenario e in stage 2

Z_p^1 : number of production line for manufacturing product type *p* in stage 1
Z_{pe}^2 : number of production line for manufacturing product type *p* under scenario *e* in stage 2
ψ: number of required workers for operating one line
W^1: Number of workers in stage 1.
W_e^2: Number of workers under scenario *e* in stage 2
V_{msr}^1: binary decision variable, $V_{msr}^1 = 1$ if material *m* is purchased from supplier *s* at price range *r* in stage 1; otherwise $V_{msr}^1 = 0$
V_{msre}^2: binary decision variable, $V_{msre}^2 = 1$ if material *m* is purchased from supplier *s* at price range *r* under scenario *e* in stage 2; otherwise $V_{msre}^2 = 0$
F_{pe}: binary decision variable, $F_{pe} = 1$ if at the end of stage 2 under scenario *e* there is no redundant product type *p*; otherwise $F_{pe} = 0$
G_{me}: binary decision variable, $G_{me} = 1$ if at the end of stage 2 under scenario *e* there is no redundant material type *m*; otherwise $G_{me} = 0$
I_{pe}: be the income which is obtained the from selling product type *p* under scenario *e* when the demand smaller than the total manufacturing quantities
O_{pe}: be the money which is obtained from selling the unsold products type *p* at savage price under scenario *e*
N_{me}: be the money which is obtained from selling the unsold materials *m* at savage price under scenario *e*

Mathematical model

The expected total cost for both two stages can be presented as:

$$TC = TC_1 + TC_2 \qquad (1)$$

where TC_1 is the total cost of stage 1 and can be computed as:

$$TC_1 = \sum_{m=1}^{M}\sum_{s=1}^{S}\sum_{r=1}^{R}\theta_{msr}^1 X_{msr}^1 + \sum_{m=1}^{M}\sum_{s=1}^{S}\sum_{r=1}^{R}\Phi_{msr}^1 V_{msr}^1 + \sum_{l=1}^{L}\omega^1 Z_p^1 + \sum_{p=1}^{P}\delta_p Y_p^1 + W^1\tau^1 \qquad (2)$$

The expected total cost TC_2 is given by:

$$TC_2 = \sum_{e=1}^{E}\pi_e\left(\sum_{m=1}^{M}\sum_{s=1}^{S}\sum_{r=1}^{R}\theta_{msre}^2 X_{msre}^2 + \sum_{s=1}^{S}\sum_{m=1}^{M}\sum_{r=1}^{R}\Phi_{msre}^2 V_{msre}^2 + \sum_{p=1}^{P}\omega_e^2 Z_{pe}^2 + \sum_{p=1}^{P}\delta_p Y_{pe}^2 + W_e^2\tau_e^2\right) - \sum_{e=1}^{E}\pi_e\left(\sum_{p=1}^{P} I_{pe} + \sum_{p=1}^{P} O_{pe} + \sum_{m=1}^{M} N_{me}\right) \qquad (3)$$

Subject to:

$$\sum_{r=1}^{R} V_{msr}^1 \le 1, \forall m = 1 \dots M, s = 1 \dots S \qquad (4)$$

$$\sum_{r=1}^{R} V_{msre}^2 \le 1, \forall m = 1 \dots M, s = 1 \dots S, r = 1 \dots R, e = 1 \dots E \qquad (5)$$

Eqs. (4-5) force the materials purchased from each supplier be either zero or belong to only one price range

$$LB_{msr}^1 \le X_{msr}^1 + BigM(1 - V_{msr}^1), \forall m = 1 \dots M, s = 1 \dots S, r = 1 \dots R \qquad (6)$$

$$X_{msr}^1 \le UB_{msr}^1 + BigM(1 - V_{msr}^1), \forall m = 1 \dots M, s = 1 \dots S, r = 1 \dots R \qquad (7)$$

$$X_{msr}^1 \le UB_{msr}^1 V_{msr}^1, \forall m = 1 \dots M, s = 1 \dots S, r = 1 \dots R \qquad (8)$$

$$LB_{msre}^2 \le X_{msre}^2 + BigM(1 - V_{msre}^2), \forall m = 1 \dots M, s = 1 \dots S, r = 1 \dots R, e = 1 \dots E \qquad (9)$$

$$X_{msre}^2 \le UB_{msre}^2 + BigM(1 - V_{msre}^2), \forall m = 1 \dots M, s = 1 \dots S, r = 1 \dots R, e = 1 \dots E \qquad (10)$$

$$X_{msre}^2 \le UB_{msre}^2 V_{msre}^2, \forall m = 1 \dots M, s = 1 \dots S, r = 1 \dots R, e = 1 \dots E \qquad (11)$$

Eqs. (6-8) state that if $V_{msr}^1 = 1$, the purchasing quantity X_{msr}^1 from supplier *s* must lie in the proper range, i.e., $LB_{msr}^1 \le X_{msr}^1 \le UB_{msr}^1$; otherwise X_{msr}^1 =0.

Eqs. (9-11) ensure that if $V_{msre}^2 = 1$, then $LB_{msre}^2 \le X_{msre}^2 \le UB_{msre}^2$; otherwise X_{msre}^2 =0

$$\sum_{s=1}^{S}\sum_{r=1}^{R} X_{msr}^1 + \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msre}^2 \le \sum_{p=1}^{P} D_p^{max}\alpha_{mp}, \forall m = 1 \dots M, e = 1 \dots E \qquad (12)$$

The total purchasing materials in both two stages must be smaller than the maximize usage of them in all scenarios are shown in Eq.(12)

$$\sum_{r=1}^{R} X_{msr}^1 \le C_{ms}^1, \forall m = 1 \dots M, s = 1 \dots S \qquad (13)$$

$$\sum_{r=1}^{R} X_{msre}^2 \le C_{mse}^2, \forall m = 1 \dots M, s = 1 \dots S, e = 1 \dots E \qquad (14)$$

Eqs. (13-14) represent the constraints purchasing material quantities due to supplier capacities for each stage

$$\sum_{p=1}^{P}\alpha_{mp} Y_p^1 \le \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msr}^1, \forall m = 1 \dots M \qquad (15)$$

$$\sum_{p=1}^{P}\alpha_{mp} Y_p^1 + \sum_{p=1}^{P}\alpha_{mp} Y_{pe}^2 \le \sum_{s=1}^{S}\sum_{r=1}^{R}(X_{msr}^1 + X_{msre}^2), \forall m = 1 \dots M, e = 1 \dots E \qquad (16)$$

Eqs. (15-16) claim that the number of product type *p* produce on each stage must be smaller than the total available of material type *m* on the stage

$$t_p Y_p^1 \leq H_p^1 Z_p^1, \forall l = 1 \dots L \quad (17)$$

$$t_p Y_{pe}^2 \leq H_{pe}^2 Z_{pe}^2, \forall p = 1 \dots P, e = 1 \dots E \quad (18)$$

$$Z_p^1 \leq \vartheta_p^1, \forall p = 1 \dots P \quad (19)$$

$$Z_{pe}^2 \leq \vartheta_{pe}^2, \forall p = 1 \dots P, e = 1 \dots E \quad (20)$$

Eqs. (17-18) force the usage of time for manufacturing product type p in each stage must be smaller than the number of available hours. Eqs. (19-20) ensure that the number of lines for each product must be smaller than its maximum value.

$$W^1 \geq \psi \sum_{p=1}^{P} Z_p^1 \quad (21)$$

$$W_e^2 \geq \psi \sum_{p=1}^{P} Z_{pe}^2, \forall e = 1 \dots E \quad (22)$$

Eqs. (21-22) create the constraints that the number of worker in each stage must greater than the minimum number of required workers for line operation.

$$O_{pe} = \beta_p^P max\left(Y_p^1 + Y_{pe}^2 - D_{pe}, 0\right), \forall p = 1 \dots P, e = 1 \dots E \quad (23)$$

$$N_{me} = \beta_m^M max\left(\sum_{s=1}^{S}\sum_{r=1}^{R} X_{ms}^1 + \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msre}^2 - \sum_{p=1}^{P}\sum_{l=1}^{L} \alpha_{mp} Y_{pl}^1 - \sum_{p=1}^{P}\sum_{l=1}^{L} \alpha_{mp} Y_{ple}^2), 0\right),$$

$$\forall p = 1 \dots m, e = 1 \dots E \quad (24)$$

Eqs. (23-24) show how the savage benefit is calculated when the company manufactured or purchased more than demand in each scenario. To linearize the max constraints in Eqs. (23-24) two variables F_{pe} and G_{me} are introduced and formulated as

$$F_{pe} = \begin{cases} 1 \ if \ Y_p^1 + Y_{pe}^2 - D_{pe} \geq 0 \\ 0 \ otherwise \end{cases} \quad (25)$$

$$G_{me} = \begin{cases} 1 \ if \ \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msr}^1 + \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msre}^2 - \sum_{p=1}^{P} \alpha_{mp} Y_p^1 - \sum_{p=1}^{P} \alpha_{mp} Y_{pe}^2 \geq 0 \\ 0 \ otherwise \end{cases} \quad (26)$$

The Eq. (25) can be rewritten under linear form as

$$\left(Y_p^1 + Y_{pe}^2 - D_{pe}\right) \geq BigM\left(F_{pe} - 1\right), \forall p = 1 \dots P, e = 1 \dots E \quad (27)$$

$$\left(Y_p^1 + Y_{pe}^2 - D_{pe}\right) \leq BigM \times F_{pe}, \forall p = 1 \dots P, e = 1 \dots E \quad (28)$$

With the Eqs. (27-28), the constraint of Eq. (23) can be expressed as

$$O_{pe} \leq BigM \times F_{pe}, \forall p = 1 \dots P, e = 1 \dots E \quad (29)$$

$$\beta_p^P\left(Y_p^1 + Y_{pe}^2 - D_{pe}\right) \leq O_{pe}, \forall p = 1 \dots P, e = 1 \dots E \quad (30)$$

$$O_{pe} \leq \beta_p^P\left(Y_p^1 + Y_{pe}^2 - D_{pe}\right) + BigM \times \left(1 - F_{pe}\right), \forall p = 1 \dots P, e = 1 \dots E \quad (31)$$

Using the same transformation approach, Eq. (24) can be linearized by Eqs. (32-36) as:

$$\beta_m^M\left(\sum_{s=1}^{S}\sum_{r=1}^{R} X_{msr}^1 + \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msre}^2 - \sum_{p=1}^{P} \alpha_{mp} Y_p^1 - \sum_{p=1}^{P} \alpha_{mp} Y_{pe}^2\right) \geq BigM(G_{me} - 1), \forall m = 1 \dots M, e = 1 \dots E \quad (32)$$

$$\beta_m^M\left(\sum_{s=1}^{S}\sum_{r=1}^{R} X_{msr}^1 + \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msre}^2 - \sum_{p=1}^{P} \alpha_{mp} Y_p^1 - \sum_{p=1}^{P} \alpha_{mp} Y_{pe}^2\right) \leq BigM \times G_{me}, \forall m = 1 \dots M, e = 1 \dots E \quad (33)$$

$$N_{me} \leq BigM \times G_{me}, \forall m = 1 \dots M, e = 1 \dots E \quad (34)$$

$$\beta_m^M\left(\sum_{s=1}^{S}\sum_{r=1}^{R} X_{msr}^1 + \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msre}^2 - \sum_{p=1}^{P} \alpha_{mp} Y_p^1 - \sum_{p=1}^{P} \alpha_{mp} Y_{pe}^2\right) \leq N_{me}, \forall m = 1 \dots M, e = 1 \dots E \quad (35)$$

$$N_{me} \leq \beta_m^M\left(\sum_{s=1}^{S}\sum_{r=1}^{R} X_{msr}^1 + \sum_{s=1}^{S}\sum_{r=1}^{R} X_{msre}^2 - \sum_{p=1}^{P} \alpha_{mp} Y_p^1 - \sum_{p=1}^{P} \alpha_{mp} Y_{pe}^2\right) + BigM \times (1 - G_{me}) \leq, \forall m = 1 \dots M, e = 1 \dots E \quad (36)$$

$$I_{pe} = \gamma_p min\left(Y_p^1 + Y_{pe}^2, D_{pe}\right), \forall p = 1 \dots P, e = 1 \dots E \quad (37)$$

Eq. (35) shows that if demand quantity is less than the manufacturing or purchasing quantity in each scenario, all products can be sold with the price γ_p. This equation can be linearized by Eqs. (36-39)

$$\gamma_p D_{pe} \geq I_{pe}, \forall p = 1 \dots P, e = 1 \dots E \quad (38)$$

$$I_{pe} \geq \gamma_p D_{pe} - BigM\left(1 - F_{pe}\right), \forall p = 1 \dots P, e = 1 \dots E \quad (39)$$

$$\gamma_p\left(Y_p^1 + Y_{pe}^2\right) \geq I_{pe}, \forall p = 1 \dots P, e = 1 \dots E \quad (40)$$

$$I_{pe} \geq \gamma_p\left(Y_p^1 + Y_{pe}^2\right) - BigM \times F_{pe}, \forall p = 1 \dots P, e = 1 \dots E \quad (41)$$

To test the validation of the model, several instances of the problem have been created. Results show that, materials owning high difference cost between the first and the second stages take more priority in purchasing activity in the first stage. As a result of that, they are often purchased for covering the usage in the whole planning horizon. These decisions are really constrast with the remained materials which have low difference cost between two stages. For these materials, decision maker will apply the delay strategy where only fraction of materials is purchased in the first stage and the remains are decided later after demands have been realized. If cost for manufacturing one product or cost for opening one line are high, manufacture often accepts lost sale. However, if the products have high selling prices, they are often manufactured in large quantities which create overage products in some scenarios.

4. Conclusion

The aim of this paper is to apply stochastic two-stage programming to some main crucial activities in supply chain such as purchasing decision and production planning in order to minimize the risk of uncertain data through considering the historical data under probabilistic term. A two-stage stochastic linear programming approach is proposed for procurement and manufacturing process, considering material all-unit discount to place optimal quantity for production. The result not only provides an average expected profit earned back for all scenario but also provides an optimal order quantity from specific supplier for both stages and the detailed production plan in each scenario. However, because of its limited number of variables that CPLEX software can solve and the complexity of the model, wide-range model more than 20 scenarios can take a long time to process. The model can perform well in small or medium range. Future research can transfer the model into a more efficient metaheuristic algorithm.

Acknowledgement: This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number C2020-28-10.

5. Reference

Aouam, T., and Uzsoy, R. (2014) “Zero-order production planning models with stochastic demand and workload- dependent lead times,” *International Journal of Production Research*, vol. **56**, no. 6, pp. 1661–1679, 2014.

Altendorfer, K., Ferberbauer, T., and Jodlbauer, H. (2016) Effects of forecast errors on optimal utilisation in aggregate production planning with stochastic customer demand, *International Journal of Production Research*, **54**, 3718-3735.

Birge, J.R., Louveaux, F. (2011). *Introduction to stochastic programming.* Springer Series in Operation Research and Finance Engineering. Springer.

Fang, C., Liu, X., Pardalos, P.M., Long, J., Bei, J. *et al*., (2017) A stochastic production planning problem in hybrid manufacturing and remanufacturing systems with resource capacity planning, *Journal of Global Optimization,* **68**, 851-878

Leung, S.C.H., Wu, Y. and Lai, K.K. (2006) A stochastic programming approach for multi-site aggregate production planning, *Journal of Operation Reserach Society,* **57**, 123-132.

Leung S.C.H. and Ng, W.-L. (2007) A stochastic programming model for production planning of perishable products with postponement, *Production Planning & Control*, **18**, 190-202.

Mahdavi, I., Taghizadeh, K., Bagherpour, M. and Solimanpur, M. (2012) Modelling of multi-period multi- product production planning considering production routes,” *International Journal of Production Research*, **50**, 1749–1766.

Mateo, J., Pla, L.M., Solsona, F., and Pages, A. (2016) A production planning model considering uncertain demand using two-stage stochastic programming in a fresh vegetable supply chain context, *Springerplus,* **5**, 1-16.

Zanjani, M.K., Nourelfath, M. and Ait-Kadi, D. (2009) A multi-stage stochastic programming approach for production planning with uncertainty in the quality of raw materials and demand, *International Journal of Production Research*, **48**, 4701-4723.

Zanjani, M.K., Nourelfath, M., and Ait-Kadi, D. (2011) A stochastic programming approach for production planning with uncertainty in the quality of raw materials: a case in sawmills, *Journal of Operation Reserach Society,* **62**, 1334-1343.

Zhang, X., Prajapati, M., and Peden, E. (2011) A stochastic production planning model under uncertain seasonal demand and market growth, *International Journal of Production Research*, 49, 1957-1975.

ENHANCEMENT OF EFFICIENCY BY IMPROVING WORKSTATION: A CASE STUDY IN GARMENT COMPANY

Truong Thanh Tam[1], Nguyen Kim Ngan[2] and Do Ngoc Hien[3]

[1] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh city Universityof Technology – Ho Chi Minh National University,*
tttam.sdh20@hcmut.edu.vn

[2] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh city University of Technology – Ho Chi Minh National University,*
nkngan.sdh19@hcmut.edu.vn

[3] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh city University of Technology – Ho Chi Minh National University,*
hienise97@hcmut.edu.vn

Abstract. Manufacturing is considered as the heart of the garment factory because it directly contributes profits to the company. In the sewing line, workers are considered as an important resource, which contribute more than 85% of the factory productivity. The remaining 15% of productivity is determined by operating method. Analysis, measurement and work standard activities are carried out every day to accurately assess the real value of the production line. This paper investigates the effectiveness of improving workstation to increase efficiency in sportswear production lines. The current state of workstations was identified by using Standardization and Work design processes (SW) with three levels: Standardized motion, standardized workstation, and improvement workstation. With these proposed methods, the existing workstations were divided into Value-adding (VA) and Non-value adding (NVA) operations. Consequently, the latter were eliminated or reduced by suitable alternatives. As a result, operation time, which might bring benefits to employees and the company, was reduced. The performance of the workstations has significantly increased it indicated that SW method was effective for analyzing workstations to solve productivity problems.

Keywords: Work Design, Standardize of Work, Lean manufacturing

1. Introduction

Standardized work (SW) is an action-oriented procedure, an essential element of the lean production system (Ohno, 1988). Besides, SW is also one of the most powerful tools in lean manufacturing that can be used to establish the best job execution sequence for each process, machine, and worker (Kasul et al. 1997). In the Toyota Production System (TPS), standardization of work is seen as important in sustaining any innovation (Jaffar et al. 2012). Standardized work creates the basis for Kaizen and if there is SW, improvement is maintained and for further improvements, so standardization is a never-ending process (MÍkva, 2016)

SW is designed by studying and observing process details, based on product and customer requirements. The main goal of the work standard is to minimize process variation between workers, eliminate unnecessary or non-value added (NVA) motion and create good quality products, safe and economics (Kasul and Motwani, 1997). SW also provides the best reference for training new workers on how to best perform processes and eliminate waste consistently, efficiently, safely, while ensuring quality, without mistakes and on-time delivery (Kasul and Motwani, 1997).

Through many literatures, most of companies are feeling the inefficiencies and loss of productivity due to poor application of ergonomics because jobs with high-energy expenditure and whole-body fatigue are the primary cause of increased costs and loss of productivity (Kaudewitz et al. 1998) so that we need to deeply understand about ergonomics and the work activity when design the workplace to all employees see the workplace is the same as their home.

Because many benefits from the standardized work contributing to any company, so there are a lot of studies in this field. However, most of them only showed the is mainly theoretical presentation of the benefits of SW to production (Fireman et al, 2018) or results of their studies rather than focusing on the detailed description of the implementation process which is useful in the current manufacturing industry (Jadhav, 2017). One of the primary related studies is the work of Halim et al. (2015) about the effectiveness of standardized work in the Air Purification Module assembly line. In their paper, they used SW as a method to find out non-value adding (NVA) operation before applying kaizen to eliminate wasteful movement. The cycle time of all motions of this process were observed and recorded by using standardize work combination table (SWCS) and Standardized Work Chart (SWC) to analyze the movements of the operator. As a result, cycle time was reduced by 16 percent which was similar to 16 percent improved productivity. This work could be considered as the first step for experimental studies in this field because its purposes and applications were constructed and derived respectively from the actual demand of production.

The aim of this article is to highlight ergonomic practices in the workstation in the garment industry by describing in detail the process of analyzing and improving workstations using a simple standardization process for design workstations. After standardized, processing time was reduced and non-value adding activities were eliminated also. This in turn can save resources to increase work efficiency and create a comfortable environment for employees.

2. Methodology

2.1. Time study

Time study is a technique that sets a time-allowed standard to perform a given task, with reasonable compensation for the inevitable fatigue, personality and delay of workers (Espinoza, 2010). Time study is used to set a standard time by observing one worker working on several cycles (Barnes, 1980).

Implementation of time study is really both an art and a science (Groover, 2007). Including 4 basic steps as below:

Step 1: A task is divided into several sub-sections and selects specific parts.

Step 2: Using a stopwatch to check cycle time for a trained worker performing those parts over a number of execution cycles, and then time each part of the work.

Step 3: Determine the sample size for the required accuracy.

To ensure reliability, the Sample should be determined by the formula:

$$n = \left(\frac{Z.S}{a.\bar{x}}\right)^2$$

Where:

z = Number of normal standard deviations needed for desired confidence

s = Sample standard deviation

a = Desired accuracy percentage (α factor)

$\bar{x}$ = Sample mean

Step 4: Set a standard time by assessing work completion and reducing time to avoid fatigue.

Normal time (NT) = $\bar{t}.(F_i).(RF_i)$

Where:

$\bar{t}$ = Time per work element

F = Frequency (number of executions in n observations)

RF = Task completion rating (working speed vs. expected average speed

If the operation has many elements: NTC = $\sum NT_i$

Standard time (ST) = NTC. (1+A), *with A is allowance to the normal time to compute the standard time.*

Workstation design (WD) technique is to create a system of work that employees need to do to make their work more effective. Workstation design has a huge impact on worker productivity. It not only makes use of the best working space through optimal and efficient machine arrangement including human integration, but also connects the workplace with its surroundings through research Ergonomics helping the workers to limit the movements that cause fatigue (Thomas, 2012). In order to work design to be productive as well as ergonomically, a research methodology was introduced that consisted of Six main steps (Do, 2014) which was showed in Figure 1 below.

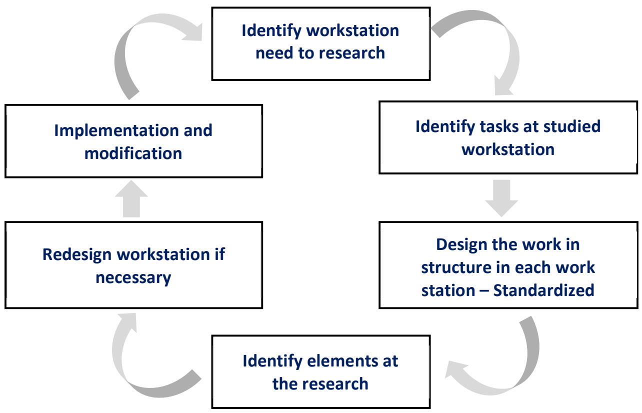


Figure 1. Cycle of Work Design process

The important recommendations of design workflow process for shop-floor production with six steps are suggested as follows:

Step 1. Identify workstation need to research

In this step, the Research Objects have to be identified. It includes the workstation and all operations in this station. The purpose of this step is to identify which workstations are experiencing problems and should prioritize improvement.

Step 2. Identify tasks at studied workstation

In order to have a better understanding of the workstation, the process flow in the workstation should be recorded; it should include all required tasks and the sequence required task to do the work. Also in this step, a few improvement ideas were created based on the human factor.

Step 3. Design the work in structure in each workstation – Standardized work

By designing the work instruction, the required sequence task to do the work was mentioned, includes all operations and work movements so that the product was produced with the best quality and the shortest time. This also is the goal of this step, a good Work instruction will help limit unnecessary activities during the job implementation, it makes the work is standard to increase worker productivity.

Step 4. Identify elements at the research

The goal of this step is to determine whether the essential and unnecessary elements in the workstation. Consider whether they are streamlined. Rejection or improvement can be reviewed if necessary.

Step 5. Redesign workstation if necessary

Depending on the actual status of the workstation, a research method will be suggesting solving the problem at this station. Additionally, the unnecessary elements are removed in this step, and the remaining essentials are relocated, put them in their place. Moreover, the redesigned tools, working table, chair or support tools can help operators do their work in the best way.

Step 6. Implementation and modification

In this step, a trial model should be made before applying in mass. Workers should be trained to better understand the purposes and benefits of the experiment. Productivity and efficiency should be measured and recorded, preferably compared with the previous status of this workstation. If the model is suitable and has significant efficiency, it should be replicated; otherwise, it should be modified to suit the actual production situation.

3. Implementation of work design process on the sportswear company: A case study

A typical research case is a sportswear production line assembled at a garment company in Vietnam. Vietnam's garment manufacturing industry is one of the major contributors to the country's GDP. Over the years, the garment industry in Vietnam has many positive changes because of applying many new production methods, especially lean production. The studied company is similar, which is in the early stages of adopting lean production tools and providing positive feedback from production. This research was applied for the sewing line with many workstations, the efficiency of this line almost based on operator's skill, thus the good workstation or well-organized work area will bring many benefits, directly affects worker productivity. In this paper, the results are presented for three types of standardization in the garment industry, such as: Workstation standardization, Motion's standardization and workstation improvement at three operations: Joining contrast panel of sleeves (OL machine), standardized sewing method: Over lock side seam at front and back body and Make buttonhole (Button hole machine). Each operation is viewed as a workstation, the results will be shown in detail as followings.

In the Joining Contrast panel of Sleeves operation, the workstation was determined including an overclock machine, operator and other necessary elements as shown on the Figure 2. This function is to join many cut-parts into the sleeves semi-product.

Figure 2. Joining Contrast panel of Sleeves Workstation

After identifying the workstation detail in the first step of the MD cycle process, all operations or tasks in the procedure are represented as a reference procedure. The operation is continuous with the combination of overclock machine, operator, and motions. Firstly, the operator will take the cut-part 1 after that put on the OL machine and running MC. the operator put the product of the task 1 to machine, get the cut-part 2 and sewing similar with task 1. The next task is repeated in the next pieces 3 and 4 for finished operation. In a garment (Polo) we have 2 sleeves so amount of required task was duplicated.

Through observation, a total of 22 motions were done by the operator to complete 1 semi-product (2 Sleeves), but by the similar of the left and right of the sleeves, this paper only concern about the process to finish 1 sleeve (Left or right). The cycle times of operations were recorded by using the stopwatch. To determine the standard workstation time, a popular used method until today is time study method which was recommended by Espinoza (2010).

Firstly, 20 samples were checked the cycle time to recalculate required the sample size for each task and based on that to determine the standard time for this operation as the formula:

Table 1. Cycle check data for the sewing 40 cm task (sec.)

7	9	8	8	8	7	9	8	9	9
8	9	7	9	8	8	9	7	8	9

Based on the cycle check in Table 1, the average of the data set ($\bar{x}$) can be determined as 8.2 (sec.) with the standard deviation (s) is 0.77. We have 20 data samples, corresponding with student distribution value (z) when reliability achieved 95% is 2.093 and the variable maximum acceptable (e) is:

$$e = a.\ \bar{x} \ = 0.05*8.2 = 0.41 \text{ (Sec.)}$$

Then, quantity samples are calculated as follows:

$$n= \left(\frac{z.s}{e}\right)^2 = 15.36 \text{ (times)}$$

So, the number of samples required for this operation is 16 (times). The remaining operations were preceded one by one; we have the Normal time data for the whole process (NTC) as following Table 2.

Table 2. The normal time of the workstation (OL 1 sleeve)

Task	Description	Normal time (Sec.)
1	Take the 1st Cut part with 2 hand and put into OL machine	3
2	Sewing 40 cm	8.2
3	Put 1st semi-product (after 3rd task) into OL machine	2.7
4	Take the 2nd cut-part then put into OL machine	3.1
5	Sewing 40 cm	8
6	Put 2nd semi-product (after the 6th task) into OL machine	2.8
7	Take the 3rd cut part then put into OL machine	3
8	Sewing 40 cm	8.2
9	Put 3rd semi-product (after the 9th task) into OL machine	2.9
10	Take the 4th cut part then put into OL machine	3.1
11	Sewing 40 cm	8
	Normal time (NTC)	**53**

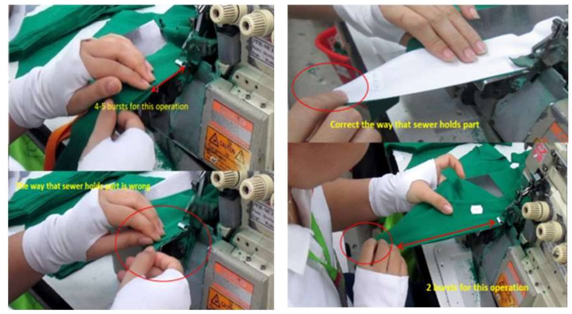
The standard time of this workstation was determined by normal time after plus 11% allowance time (it includes machine delay - 2% and the basic operator fatigue time – 9%) with the formula in Table 3 below:

Table 3. Standard time (ST) workstation OL Sleeves

Task	Description	The normal time	
		Before improved	**After improved**
1	Take the 1st Cut part with 2 hand and put into OL machine	3	2.5
2	Sewing 40 cm	8.2	8.2
3	Put 1st semi-product (after 3rd task) into OL machine	2.7	2.7
4	Take the 2nd cut-part then put into OL machine	3.1	2.8
5	Sewing 40 cm	8	8
6	Put 2nd semi-product (after the 6th task) into OL machine	2.8	2.8
7	Take the 3rd cut part then put into OL machine	3	2.4
8	Sewing 40 cm	8.2	8.2
9	Put 3rd semi-product (after the 9th task) into OL machine	2.9	2.9
10	Take the 4th cut part then put into OL machine	3.1	2.5
11	Sewing 40 cm	8	8
	Normal time (NTC)	**53**	**51**
	Allowance time (A) (11%)	**5.83**	**5.61**
	Standard time (ST)	**58.83**	**56.61**

In spite of using the small materials and a simple support tool, they can bring out much effectiveness at the workstation with the waste time be eliminated (2.22 seconds /products). We can see that the cycle time was reduced from 58.83 seconds to 56.61 seconds and productivity increased more than 4 percent.

In the manufacturing process, materials were processed by hand typically account for 35-40% of the operating time. Standardization is necessary because of its usefulness in eliminating redundant movements. In the apparel industry, movements are coded with a code, called GSD, and the operation time is calculated as the total time of necessary movements to provide value of the activity. The application of GSD time was considered to increase specialization in production. This was also a premise to boost productivity of workers because the closer the operation time is to GSD time, the more standardized and working efficiency is higher. In order to analyze operator movements, the second level of the standardized workstation is defined as the standardized sewing treatment method at station 02 (Overlock side stitching on front and rear bodies). This is a popular method to reduce wasted production time. Figure 5 illustrates standardization of work before (A) and after the kaizen process (B).



(A) Before kaizen (B) After kaizen

Figure 5. Standardized sewing handling method

Before applying kaizen, employees will handle products in 4 clusters to align cut-parts before performing next steps with a total processing time of 40 seconds. Movements that do not increase in value are determined after analyzing the workstation and those movements should be improved. By keeping the cut parts precisely, the operator can process product sales with 2 bursts as well as the processing time is reduced to 31 seconds per piece. As a result, more than 22% of the workstation's performance was increased.

By implementing kaizen in production, the efficiency of operations has been improved by reducing manual workloads, increasing automatic workload ratios with the help of tools and machines. After being standardized, repetitive value-adding activities are replaced with. In this case, kaizen model is proposed to replace the manual operations of the punching machine (Figure 6).

Figure 6. Button hole operation (Before)

The current process, the worker spent 17 seconds for 4 stages: the first, worker will take and put the lining piece into buttonhole machine after that load the semi-product into machine, press the button to running machine and loading out the semi-product in the four steps.

The repeated activity was identifying after observation, this is loading lining into machine before loading semi-product. Several ideas for reducing the cycle time have been proposed. The last idea was selected by its benefit, with the workstation is 15.6 seconds (reduced by 1.4 seconds per product). The results before and after performing the kaizen process are shown in Figure 7:

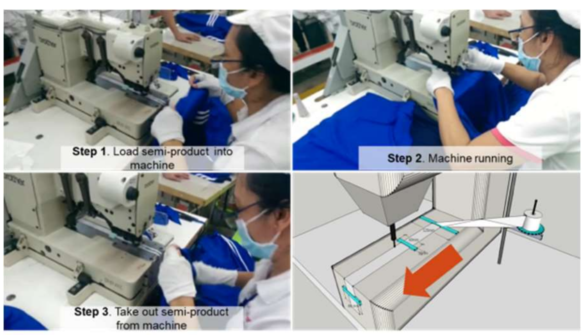


Figure 7. Buttonhole operation (After kaizen)

Observing the results when applying the kaizen method, we see that the productivity of the workstation increases by more than 8.2%.

4. Result and discussion

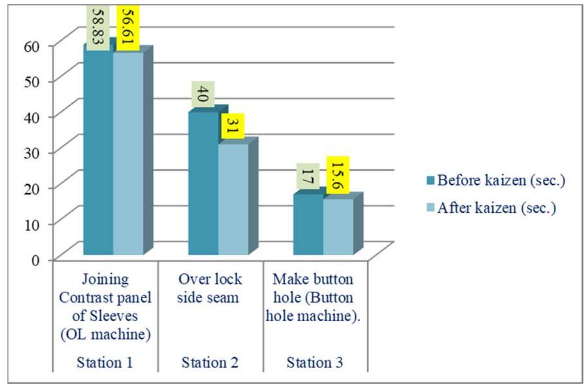


Figure 8. Summarize of standardized workstation result.

The efficiency of the production line depends not only on the skills of the workers, but also on the appropriate implementation method to minimize the generation of waste. Through standardization, workstations have been redesigned for better workplaces. The aim is to reduce the job cycle time as well as reduce fatigue to create a friendly working environment for employees, thereby improving the productivity of the workstation. The data was shown in Figure 8.

Through analysis and evaluation, the performance of workstations has been significantly improved: 2% in the first station, 22% in the second station and 8.2% in the third station. Optimizing resource efficiency can help reduce costs and resources for production

5. Conclusions

In this research, value-added and non-value-added activities have been considered through the application of the building design model mentioned in section 2. The model includes the steps that are assumed to be needed to determine the value-adding and non-value-adding operations to eliminate or limit the second operation. The steps in the model are presented clearly and systematically, from data collection method, data processing to execution on sportswear production line. Applying work redesign has helped to reduce working time thereby increasing employee productivity, reducing costs and bringing profit to the company. The results of the study can be widely researched and developed in the garment industry as the basis for analysis and implementation of kaizen to reduce cycle time as well as increase production efficiency for the company.

As it is one of the most powerful tools in lean manufacturing, it should be used in conjunction with other tools to get the best results for the sewing line. Besides, to apply successfully, it requires support and commitment of top managers as well as training for more understanding. As a result, better production will be achieved.

References

Do, N. H. (2014). Standardization of the Work Design Process in the Manufacturing Shopfloor, *Conference on Mechanical and Manufacturing Engineering*

Espinoza, O. (2010). *Method, standard and work design. New York: Mc Graw–Hill.*

Fireman, M.C.T, Saurin, T.A. and Formoso, C.T. (2018). The role of Slack in Standardized Work in construction: An Exploratory Study, *26th Annual Conference of the International Group for Lean Construction (IGLC),* González, V.A. (ed.), Chennai, India, 1313–1322.

Groover, M. P. (2007). Work Systems and Methods, Measurement, and Management of Work, *Pearson Education International.*

Halim, N et al. (2015). *Jurnal Teknologi (Sciences & Engineering),* 76(6), pp. 73–78

Jadhav, S. S, Sharma, G. S., Daberao, A. M. and Gulhan, S. S. (2017). Improving Productivity of Garment Industry with Time Study. *International Journal on Textile Engineering and Processes,* 3, pp.1

Jaffar, A., Halim, N. H. A. and Yusoff, N. (2012). Effective Data Collection and Analysis for Efficient Implement of Standardized Work (SW). *J. of Mechanical Engineering*, 9, 45-78

Kasul, R. A. and Motwani J. G. (1997). Successful Implementation of TPS in A Manufacturing Setting: A Case Study. *Industrial Management & Data Systems*, 97(7), pp. 274- 279.

Kaudewitz, H. R. (1998). Work standard assessment using heart rate monitoring. *IIE Solutions*, 30(9), pp.36

Liker, J. K. (2004). *The Toyota way: 14 management principles from the world's greatest manufacturer*. *New York: McGraw-Hill*. ISBN 0-07-139231-9. OCLC 54005437

MÍkva, M. et al. (2016). Standardization - one of the tools of continuous improvement. *International Conference on Manufacturing Engineering and Materials*, ICMEM 2016, Nový Smokovec, Slovakia

Monden, Y. (1983). Toyota Production system – An Integrated approach to Just in Time, *4th ed., International Standard (Book)*, 13, 978-1-4665-0451-6 (eBook – PDF)

Monk, Ellen and Bret Wagner (2009). Concepts in Enterprise Resource Planning-3rd Edition. *Boston: Course Technology Cengage Learning*.

Ohno (1988) *Workplace Management*

Tapping D, Luyster T and Shuker T. (2002). Value Stream Management - Eight Steps to Planning, *Mapping and Sustaining Lean Improvements*. New York: Productivity, Inc.

Thomas, J. (2012). Ergonomics and Workplace Design [Online]. Available at: https://www.workdesign.com/2012/08/ergonomics-and-workplace-design/

Cross Entropy Method and Genetic Algorithm approaches for the Job Shop Problem

Tran Ngoc Quynh[1,*], Do Ngoc Hien[1] and Tran Quoc Cong[2]

[1] *Department of Industrial Systems Engineering, Ho Chi Minh City University of Technology, Vietnam National University, Viet Nam*

[2] *Department of Civil and Natural Resources Engineering, University of Canterbury, New Zealand*

Abstract. The development of the optimization method has led to many new methods to solve job shop scheduling problems (JSP). This paper offers the Cross-Entropy method (CEM) and Genetic Algorithm (GA) to solve this problem, minimizing the order's max lateness. A matrix encoding scheme is proposed based on the CEM, and a backward decoding method was used to generate a reasonable schedule. To describe the distribution of the solution space, a probability distribution model is built and used to generate individuals. In addition, the probability updating mechanism of the probability distribution model is proposed, which helps to find the optimal individual gradually. The experiment results show that, generally, CEM produced competitive solutions compared to the GA.

Keywords: Job Shop Scheduling, Cross-Entropy Method, Genetic Algorithm

1. Introduction

In a typical job shop problem (JSP), each job has its unique operation route. Because the continuity of operation in each job must be kept avoiding operation reworking or job redoing, incorrect scheduling methods may make the makespan or tardiness significantly longer. Therefore, a suitable production scheduling method can ensure that the work is done with the highest efficiency with the lowest cost and help the system earn high profits.

Many scheduling algorithms have been applied in JSP, and each has its pros and cons. In general, there are three groups of scheduling algorithms: Dispatching rules, the Approximation algorithm and the Exact Optimization algorithm. According to Zhang et al. (2019), relative algorithms will work better for complex production systems such as the Job shop production system. In one study, Mattfeld and Bierwirth (2004) showed that the Genetic algorithm (GA) with an algorithm that reduces the search space through iterations could find the optimal solution in a short time.

However, in another study, Santosa et al. (2011) showed that the application of a new algorithm - the Cross-Entropy Method (CEM) works well in moderation problems and reduces calculation time compared to other commonly used algorithms like Genetic or Simulated Annealing. Therefore, this study will select the Genetic and Cross-Entropy algorithms to build a model of applying two algorithms to the JSP, with the objective is to minimize the Max Tardiness and compare the effectiveness of the two algorithms.

2. Problem description

2.1 Job Shop Problem (JSP) and Flexible Job Shop Problem (FJSP)

The Job Shop problem (JSP) consists of a set of jobs {Ji}: 1≤i≤n that need to be processed on a set of independent machines {Mr} 1≤r≤m. Each job has a path through its process sequence. The two critical constraints of the JSP are the fixed workflow on a set of machines and the limited capacity of the machines (each machine can only process one job at one time).

The Flexible Job Shop problem (FJSP) is a combination of the Job Shop problem and the parallel machine environment. Instead of a set of machines, the FJSP problem consists of a set of work centres, and each centre consists of a number of parallel machines. FJSP can be described one some points:

- The problem consists of n jobs, and each job has some o operations.
- There are m machines, and each of them can only operate one operation at a time.
- Each operation will be proceeding on a specific machine and a particular operation time.
- No breakdown, lost material, etc., will stop the production process.

* Corresponding author

The objective of the FJSP is often varied. In this study, minimizing the maximum tardiness is chosen as the problem's objective.

2.2 Cross-Entropy Method (CEM)

First proposed in 1997 by Reuven Rubinstein (1938-2012), Cross-Entropy Method (CEM) is an effective method to estimate the probability of a rare event. CEM has rapidly evolved into a powerful and flexible technique for rare event simulation and combinational optimization (Rubinstein and Kroese, 2013). The method has been successfully used in many fields of engineering and science, various problems such as locating problems (Chan et al. 2020), DNA sequence alignment (Keith and Kroese, 2002), vehicle routing problems (Chepuri and Homem-De-Mello, 2005).

Pieter - Tjerk et al. (2005) have introduced the concept and application of CEM in the optimization problem. For instance, CEM is based on sample distribution information. This information helps the algorithm capture the distribution of a good sample. The new sample will be created based on this distribution. The distribution will be updated continually through iterations until finding an optimal solution. Unlike other algorithms, CEM needs to update the probability distribution parameters of the selected elite sample after each iteration and must complete two main steps:

- Generate samples based on a defined probability distribution.
- Update its parameters based on the best sample data (elite sample). So, the next iteration will have a good sample based on this parameter.

Besides, according to Pieter – Tjerk et al. (2005), the meaning of the CEM is that it can create an accurate mathematical framework to start, thereby updating with some optimal rules based on advanced simulation theory to find the solution.

2.3 Genetic Algorithm (GA)

Genetic Algorithms (GA) was invented based on the development of natural selection and genetic ideas. This is the primary optimization algorithm using natural evolution techniques such as inheritance, mutation, selection and crossover. The father of the original Genetic Algorithm was John Holland, who invented it in 1970, and Charles Darwin introduced a random search method in a defined searching space to solve the problem [9].

In one study, Pezzella et al. (2008) showed a significant difference between GA and other methods, that GA maintains and processes a set of solutions, called population. In GA, the search for an appropriate hypothesis begins with a population or an initial selective set of hypotheses. Individuals of the current population originate the next generation population utilizing random crossover and mutation activities, sampled after biological evolution processes. At each step, individuals which are more developing and adaptive to the environment will survive, and the others will be eliminated. Therefore, GA can detect new generations with better adaptability.

GA solves optimization problems through basic processes: crossover, mutation and selection for individuals in the population. Therefore, using GA requires determination: initial population, the objective function evaluating the solutions according to the fitness, the genetic operators create the reproductive function.

3. Problem formulation

The notations used in this paper are summarized as in Table 1:

Table 1. Mathematical notation

Symbols	Definition
m	Number of machines
n	Number of jobs
i	Index of jobs, i = 1..n
j	Index of machines, j = 1..m
C_i	Completion time of job i
ts_{ij}	Start time of job i on machine j
te_{ij}	End time of job i on machine j
p_{ij}	Processing time of job i on machine j
d_i	Due date of job i

The mathematical model is proposed to minimize the maximum tardiness of jobs; with decision variables and problem formulation are written as below:

$$Minimize\ T_{max} \quad (1)$$

S.t:

$$T_{max} = \max\{T_i\}; \quad \forall i \in [1, n] \quad (2)$$

$$T_i = \max(C_i - d_i, 0); \ \forall i \in [1, n] \quad (3)$$

$$C_i = \max\{te_{ij}\}; \quad \forall i \in [1, n] \quad (4)$$

$$te_{ij} = ts_{ij} + p_{ij};\ \forall i \in [1,n];\ j \in [1,m] \quad (5)$$

$$ts_{ij} \geq te_{j'};\quad \text{j' is the preceding operation of j} \quad (6)$$

$$ts_{ij}, te_{ij} \geq 0;\ \forall i \in [1,n];\ j \in [1,m] \quad (7)$$

Equation (2) describes the maximum lateness of jobs, which is equal to the difference between completion time and the due date of a job and take the positive number as in constraint (3). Completion time of job i is the end time of job i's final operation (4). Furthermore, the end time of processing job i on machine j is equal to the sum of start time and processing time of processing job i on machine j (5). Besides, constraint (6) guarantees the order of operation route and constraint (7) is the constraint of non-negative binding of the start and end times of the operation.

4. Genetic algorithm

The proposed GA for FJSP is given in the figure below:

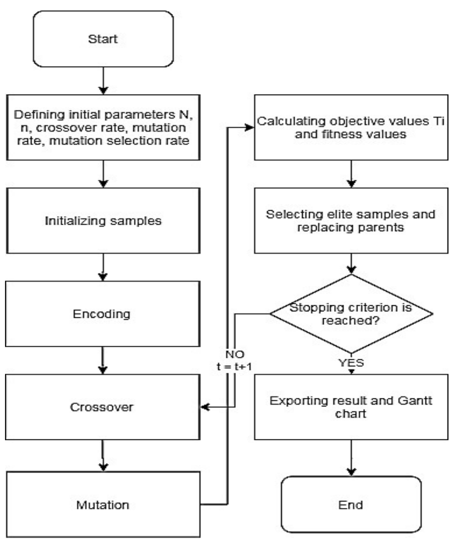


Figure 1. Flow chart of GA (Yadav and Prajapati, 2012).

For instance, at the stage of defining input parameters, it is necessary to determine sample size N, crossover rate, mutation rate, selection rate and several iterations n. After consulting related studies, the algorithm's input parameters are selected as follows:

- Sample size N: since the sample size does not have a certain threshold, the larger the number of jobs, the larger the sample size is required due to the permutation in the large dispatch. In this study, choose sample size N = 200
- The crossover rate is equal to 0.8
- The mutation rate is equal to 0.2
- The selection rate is equal to 0.2

The number of iterations n: To determine the influence of the number of iterations on the results and time of performing the problem, running the sample problem with different numbers of iterations and evaluating the results.

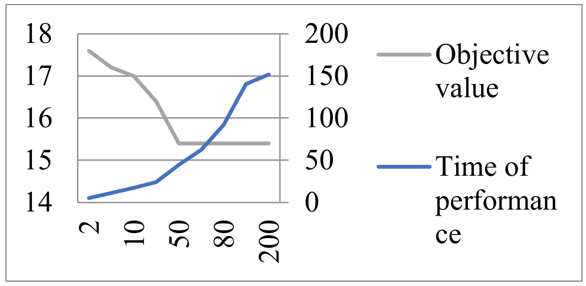


Figure 2. Influence of the number of iterations on the results and time of performing

The result in figure 2 shows that when the number of iterations increased, the objective value improved. However, the calculation time will also increase proportionally to the number of iterations. With n = 200 iterations, an optimal objective value has been obtained, with runtime at an acceptable range. Therefore, to increase the accuracy of the solution, select the number of iterations n = 200.

Next, create N random initial samples, each representing a sequence of n jobs. The initial sample sequences are then coded according to Kacem et al. (2002), each task formed by a trio (i, j, k):

- i is the notation of the job of that task
- j is the index of tasks of job i
- k is the notation of the machine performing that task

Generate offspring from "father" and "mother" using two points – crossover, two crossover points are picked randomly. After that, mutate the offspring by permutating pairs of crossed genes. Next, calculate the objective value T_{max} with a simple moderation method, and the fitness value is calculated by the inverse of the objective value. The elite samples are then selected according to the initial selection rate and replaced the samples with poor objective value in the initial population. Then, check the stopping condition: if the number of iterations n = 200 is reached, the scheduling sequence with the optimal objective value and the corresponding Gantt chart will be exported. On the contrary, continue to perform crossover and mutation for the sample set that has been replaced.

5. Cross entropy method

CEM application to the optimization problem is described as follows: Let X is a finite set of states, S is the performance function on X. To find the maximum of S over X and the corresponding state(s) at which this maximum is attained. Let us denote the maximum by γ^*. Thus,

$$S(x^*) = \gamma^* = \max S(x) \ (x \in X)$$

The process of applying CEM in scheduling, as well as an optimization problem, is as follows:

1. Begin with the matrix of the probability of the initial solution $\widehat{p_0} = (\frac{1}{n}, \dots, \frac{1}{n})$. Set t = 1
2. From the initial probability matrix, initialize the initial set of solutions and calculate the objective values of each sample.

Call $\widehat{\gamma_t}$ is the worst of ρ × 100% of the best objective value of the solution set.

3. With ρ × 100% best samples found, the probability matrix is updated according to the following formula:

$$\hat{p}_{t,j} = \frac{\sum_{i=1}^{N} I_{\{S(X_i) \geq \widehat{\gamma}_t\}} I_{\{X_{ij}=1\}}}{\sum_{i=1}^{N} I_{\{S(X_i) \geq \widehat{\gamma}_t\}}}$$

With j = 1, …, n; $X_i = (X_{i1}, \dots, X_{in})$ and t increased by 1.

4. After updating the probability matrix, if the stopping condition is satisfied, the optimal solution will be equal to the solution with the best objective value. Otherwise, increase the number of iterations t = t + 1 and return to step 2.

For instance, the proposed CEM is described as below:

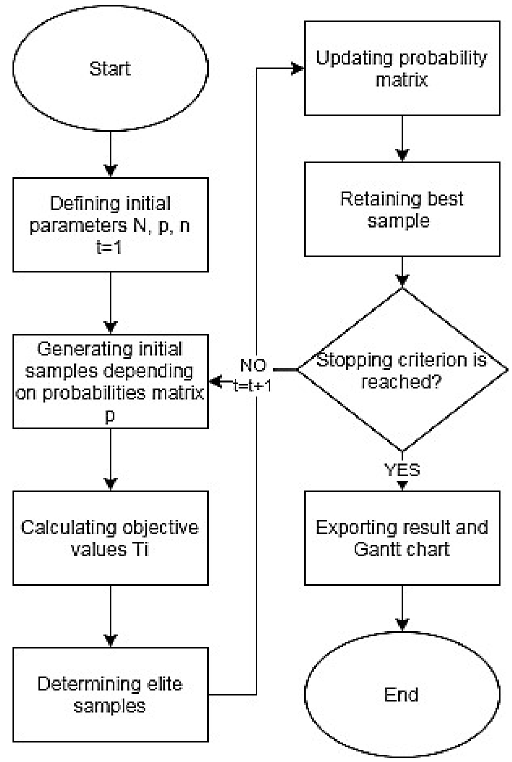


Figure 3. Flow chart of CEM

At the stage of defining input parameters, it is necessary to determine:

- Sample size N: since the sample size does not have a certain threshold, the larger the number of jobs, the larger the sample size is required due to the permutation in the large dispatch. In this study, choose sample size N = 200
- The elite sample selection rate is usually selected from 1% - 10%. Here, select ρ = 2%
- Smoothing coefficient β = 0.8

At the initializing step, each sample represents the sequence of n jobs, randomly generated from the initial probability distribution:

$$p = \begin{pmatrix} p(1,1) & \cdots & p(1,n) \\ \vdots & \ddots & \vdots \\ p(n,1) & \cdots & p(n,n) \end{pmatrix}$$

With $p(i,j) \neq 0$ corresponds to the arrangement of the job i to the place j.

Next, the calculation of the objective value T_{max} will be conducted with a simple moderation method. Then, select [ρN] samples with the best objective value from the original N samples. From the elite samples, the probability matrix is updated according to formula (4.9). Continue to initialize based on the new probability and calculate the objective value until the stopping condition is reached. In this study, the stopping condition is when ten consecutive objective values remain unchanged.

6. Numerical experiments

6.1 Input

The two algorithms were coded using Python language. The experiment is conducted with five machines x eight jobs FJSP. The input data include due dates, machine sequences and processing times, which are in Table 2, 3 and 4:

Table 2. Input data

Order	Job	Notation	Due date
0	J1	[0,1,2,3,4]	5
1	J2	[5,6,7,8,9]	6
2	J3	[10,11,12,13,14]	7
3	J4	[15,16,17,18,19]	8
4	J5	[20,21,22,23,24]	9
5	J6	[25,26,27,28,29]	10
6	J7	[30,31,32,33,34]	11
7	J8	[35,36,37,38,39]	12

Table 3. Machine sequence

Job \ Operation	O1	O2	O3	O4	O5
J1	M1	M2	M3	M4	M5
J2	M2	M3	M1	M5	M4
J3	M3	M1	M2	M5	M4
J4	M5	M4	M3	M2	M1
J5	M4	M1	M5	M3	M2
J6	M3	M5	M1	M2	M4
J7	M5	M3	M2	M4	M1
J8	M1	M4	M5	M2	M3

Table 4. Processing time

Job \ Operation	O1	O2	O3	O4	O5
J1	5	3	2	4	6
J2	2	1	5	6	4
J3	3	4	2	7	1
J4	2	8	5	1	3
J5	1	6	2	3	4
J6	5	2	4	9	3
J7	5	2	4	6	3
J8	8	5	4	3	1

6.2 GA's result

The GA's scheduling sequence of an 8x5 problem is a sequence of numbers from 0 to 7, and each number appears five times. Each number is a notation of a job, corresponding to machining in the input data table. Specifically, the results running by the program include the time to complete the job, the order of dispatch, the objective function, and the calculation time as follows:

- Makespan = 60
- Scheduling order: [0, 0, 2, 5, 2, 4, 1, 3, 1, 7, 2, 5, 6, 3, 1, 5, 3, 0, 7, 2, 4, 6, 7, 0, 3, 1, 6, 7, 5, 6, 4, 7, 4, 1, 0, 3, 5, 2, 4, 6]
- Objective value: $MinT_{max} = 35$
- Calculating time: 67.8 s

The result is exported as a sequence of jobs. Therefore, based on the machine sequence in the input table, the sequence of moderation is decoded as in Figure 4:

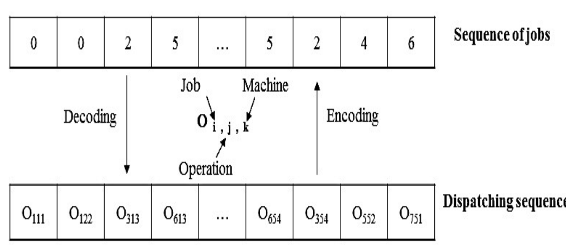


Figure 4. Encoding and decoding of GA's result

In addition, the software also supports exporting the chart showing the improvement of the solution and Gantt chart of the results as shown in Figures 5 and 6:

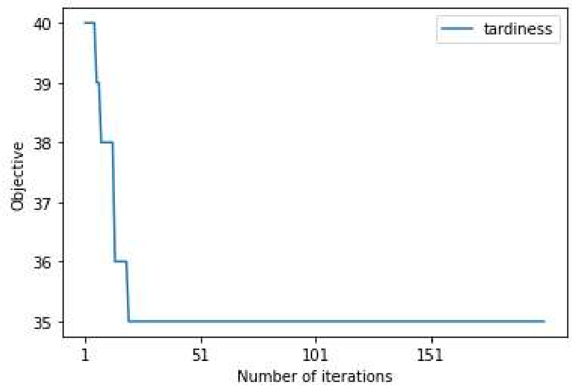


Figure 5. The improvement of the solutions

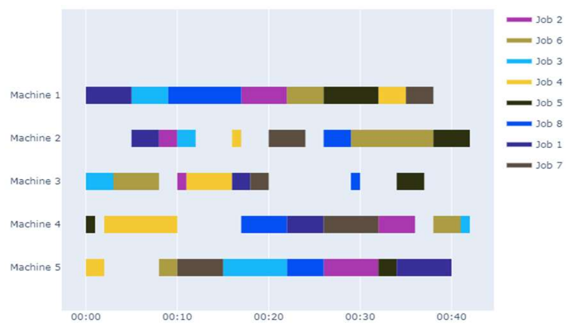


Figure 6. Gantt chart of the optimal solution

6.3 CEM's result

The CEM's result is a sequence of 8x5 = 40 numbers from 0 to 39. Since this is the problem of 8 jobs x 5 machines, every five numbers will denote the tasks of one job in the dispatching sequence. Precisely, the numbers 0, 1, 2, 3, 4 symbolize the tasks of job 1; the numbers 5, 6, 7, 8, and 9 symbolize the tasks of job 2. and the numbers 36, 37, 38, 39 symbolize the tasks of job 8, corresponding to the machining order in the input data table. The result of the problem after running by the program is as follows:

- Makespan = 45
- Scheduling order: [23. 16. 25. 4. 15. 12. 8. 1. 22. 14. 5. 34. 7. 3. 28. 30. 21. 11. 0. 13. 19. 24. 31. 6. 2. 10. 32. 36. 17. 29. 37. 20. 39. 27. 18. 33. 38. 35. 9. 26.]
- Objective value: $MinT_{max} = 34$
- Calculating time: 52.5 s

Unlike GA, CEM produces a sequence of operations. Therefore, it is necessary to decode it into a sequence of jobs before decoding into the final sequence of dispatching, as shown in Figure 7:

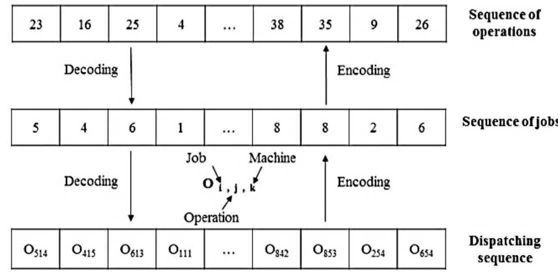


Figure 7. Encoding and decoding of CEM's result

Similar to GA, the software also supports exporting charts showing the improvement of the solution and a Gantt chart as shown in Figures 8-9:

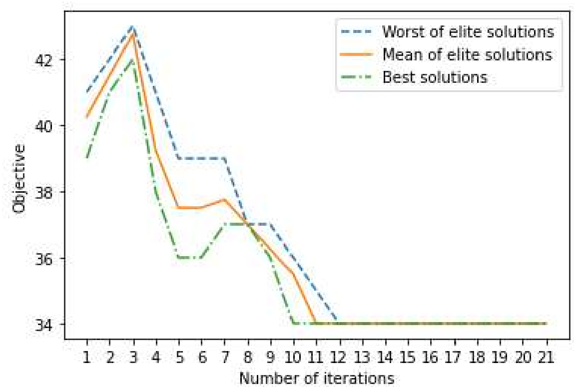


Figure 8. The improvement of the solutions

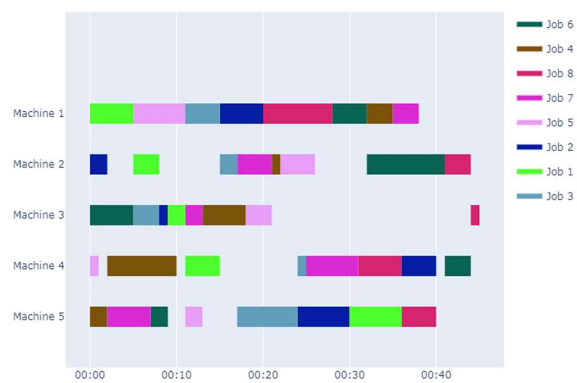


Figure 9. Gantt chart of the optimal solution

7. Analyze

The results of the 8x5 scheduling problem of 2 algorithms GA and CEM are summarized and compared in Figure 10:

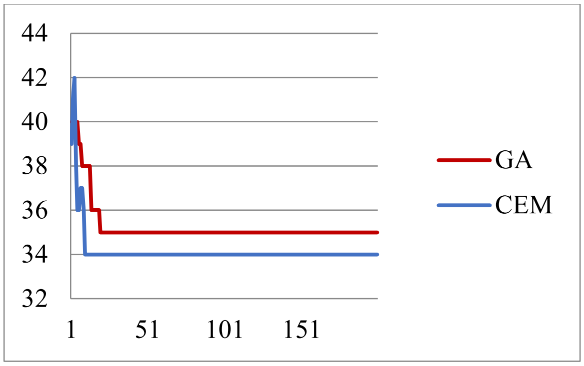


Figure 10. Results comparison

It can be seen that CEM has a better objective value than GA. Besides, the solution convergence of CEM is also better when it is possible to find an optimized solution on the 10th iteration. Meanwhile, GA needs to go through 20 iterations to find the optimal solution.

In addition, to validate the performance and compare the superiority of CEM and GA, six groups of data are solved. Sizes of 6 problems are shown in Table 5, and the results are summarized in Table 6:

Table 5. Size of sample problems

Problem	No. of machines	No. of jobs
1	3	3
2	5	8
3	10	106
4	10	35
5	10	29
6	10	69

Table 6. Summary result

Problem	**CEM**			**GA**		
	T_{max}	**C_{max}**	**Cal. time**	**T_{max}**	**C_{max}**	**Cal. ime**
1	6	13	56	6	13	50
2	35	41	608	33	66	48
3	0	6245	1905	0	5981	2236
4	0	3303	272	0	3467	499
5	953	8153	1709	1029	11090	683
6	0	6992	1296	0	7215	1241
Avg	**166**	**4125**	**974**	**178**	**4639**	**793**

The results showed that the CEM algorithm brings a better solution than GA when the average objective value T_{max} is 12 minutes smaller. Besides, the order completion time C_{max} of CEM is 514 minutes smaller than GAs. However, when considering computational time, the implementation of scheduling by CEM takes 181 seconds longer than GA. Therefore, depending on the company's needs and actual situation, appropriate algorithms can be considered to apply.

8. Conclusion

Based on the experiment results and analysis above, it is proved that CEM is an efficient method in solving JSP. It produces competitive solutions, especially in small-scale problems. CEM is much better in doing diversification solutions than GA when operating based on probability distributions. However, CEM is worse in performing large–scale problems because of the extended computing time.

In future research, CEM for JSP must be modified to get better performance, especially for large-sized and multi-objective instances. The implementation using a lower-level programming language might improve the performance of CEM.

Reference

Keith, J. and Kroese. D. P. (2002) Sequence alignment by rare event simulation.*Proceedings of the Winter Simulation Conference,* IEEE, 1

Mattfeld, D. C. and Bierwirth, C. (2004) An efficient genetic algorithm for job shop scheduling with tardiness objectives. *European journal of operational research,* 155(3), 616-630.

Rubinstein, R.Y. and Kroese, D. P. (2013) The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning. *Springer Science & Business Media.*

Santosa, B., Muhammad, A. B. and Stefanus, E. W. (2011) A cross-entropy - genetic algorithm for m-machines no-wait job - shop scheduling problem.*Journal of Intelligent Learning Systems and Applications,* 3(3), 171.

Tran, C. Q., Ngo, D. D., Keyvan-Ekbatani, M. and Watling, D. (2020) A user equilibrium - based fast-charging location model considering heterogeneous vehicles in urban networks. *Transportmetrica A: Transport Science.*

Zhang, J. et al. (2019) Review of job shop scheduling research and its new perspectives under Industry 4.0. *Journal of Intelligent Manufacturing,* 30(40), 1809-1830.

DOES LEAN ITSELF GAIN SUSTAINABLE COMPETITIVE ADVANTAGES TO MANUFACTURING?

Ton Nguyen Trong Hien[1]

[1] *School of Business Administration, Van Lang University, hien.tnt@vlu.edu.vn*

Abstract. Lean and the term 'sustainability' overlap in their intent. Traditionally there have been two central views to sustainability: Lean manufacturing as a source of sustainable competitive advantage, and Lean supports sustainability indicators in tri-pillars: economic, social, and environmental sustainability. Many studies explore the effect of Lean manufacturing practice on environmental, social, and financial to companies; however, there has been still a lack of clear and adequate measure Lean-sustainable competitive advantages. This study aims to examine whether Lean create sustainable competitive advantages dimension related to sustainability via the citation context analysis method. The claim that no one could have guaranteed Lean is enough to ensure sustainability.

Keywords: Lean production; Lean thinking; Competitive advantages; Sustainability.

1. Introduction

The term Lean was first introduced by Krafcik in 1988, but the story of Lean begins when the book 'The Machine That Changed the World' highlighted Lean, accordingly Womack and Jones (2004) describe Lean as doing "more and more with less and less" (Kaufmann, 2020). However, the Toyota system seems to "add waste rather than eliminate it" (Liker, 2003). It is not simply waste reduction; creating Lean is more work than talking about removing waste, so we also look at value-adding activities under Lean philosophy (Holweg, 2006).

Sustainable or sustainability refers to 3 primary responsibilities: economic, social, and environmental or we call the "Triple Bottom Line (TBL)" framework. Triple Bottom Line is how we produce goods and services with restricting or no pollution, energy conservation, natural resources, feasible cost, safe and healthy working environment, communities and consumers's responsibility (Krajnc and Glavic, 2005). For many years, researchers have been exploring the adapting of Lean practices to address more sustainable objectives. There is a positive impact of Lean and sustainable manufacturing on three bottom-line performances, however, there is a lack of clear and adequate measures for Lean and sustainability of manufacturing operation (Hartini and Ciptomulyono, 2015; Resta et al., 2016). Indeed, Lean has proven its positive and negative effect on three triple bottom line pillars of sustainable performance (Henao et al., 2019), thus Lean and its effect on sustainable performance is ambiguous (Iranmanesh et al., 2019).

We come to a temporary conclusion that Lean practices have not reached the maturity level required to ensure sustainability yet (Resta et al., 2016; Abualfaraa et al., 2020). Trade-offs between the conflicting aspects of corporate sustainability, plus manufacturers are urged to become more proactive regarding their environmental issue, therefore, "almost all research linking Lean operations … to sustainability issues have focused exclusively on environmental impact" (Piercy and Rich, 2015) to find out Lean practices and their tools concerning support greener production.

Alternatively, when it comes to sustainability people think of the long-term success of production systems (Liyanage 2007) which can have a huge impact on competitive advantages. In 2004, Thomas and Pham (2004) propose FIT manufacturing concept for an integrated approach to the use of leanness, agility, and sustainability to achieve the "longevity of a manufacturing organization's competitiveness". Vinodh, et al. (2014) also stated sustainability in this context is the ability of a manufacturer actively to seek to find new customers and markets, balance resources, and add product diversification. It is not much of a paradigm change, one that has not progressed sufficiently, and there would not be a mutual agreement.

It is thought that "Lean manufacturing proves competitive advantages" (Pirraglia et al., 2009), and the principal purpose of the Toyota production system is to enable organizations to respond rapidly to changing business conditions by effectively reaching the main dimensions of competitiveness: flexibility, cost, quality, service, and innovation (Shingo, 1996). The question is whether Lean itself gains a competitive advantage?

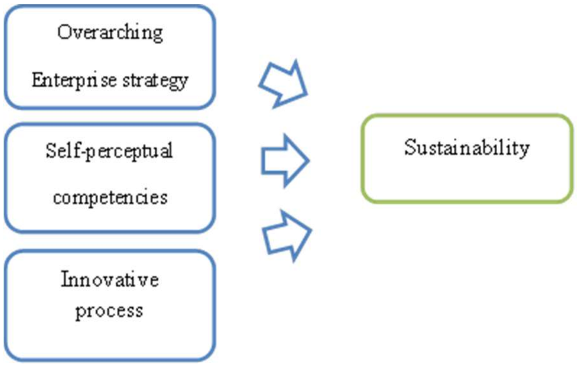


Figure 1. Component of sustainability according to FIT manufacturing perspective.

2. Research Method

To answer the research objective, a well-conducted state of the art of research into the links between Lean and sustainable competitive advantages should be given. Citation context analysis is that area of bibliometric which deals with the study of these relationships. This study uses the citation context analysis method, predominantly, it is assumed to reflect a relationship between a part of the whole of the cited document and a part of the whole of the citing document (Smith, 1981).

Stage 1: Data selection

To find articles discussion around the topic of Lean and sustainability, the analysis uses the Scopus database, which is a large, relatively reliable database for citation context analysis. The keyword search queries are specified as follows:
Key term: (TITLE (Lean) AND TITLE (sustainable) OR TITLE (sustainability) AND TITLE (production) OR TITLE (manufacturing) AND (LIMIT-TO (DOCTYPE, “ar”) AND (LIMIT-TO (LANGUAGE, “English”))

Documents were selected according to criteria:

Table 1. Target article

Inclusion criteria	Exclusion criteria
Peer-reviewed articles from Scopus	Conference paper, Book chapter, Note, v.v.
Extraction date: 12/4/2021	Excluding articles published after 12/4/2021
English articles	Excluding articles written in languages other than English

Stage 2: Observation and analysis of documents

The findings yielded a total of 67 hits of discrete articles that matched the keyword search criteria. Preliminary analysis of the returned results shows that the research topic on Lean and sustainability increased steadily in recent years, accordingly, the number of studies on this topic in 2020 is more than 1.5 times the number of studies with the same topic in 2019, and more than 8.5 times numerous studies in 2010.

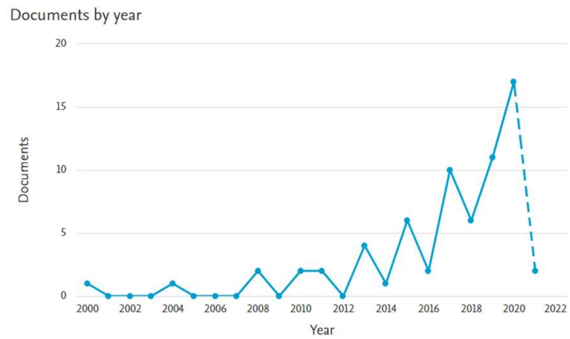


Figure 2. Chronological distribution of articles (past-12/4/2021)

The articles are presented in descending total citation count, and we found that the 1st article with an outstanding number of citations is relevant to the research topic. The 1st post had a large number of distinct citations, 1.67 and 1.92 times higher than the article with the 2nd and 3rd citations, respectively. Therefore, we decided to choose this paper for analysis and comparison.

Table 2. Focal article set

ID	Title	Cited by
1	Lean production and sustainable competitive advantage	301
2	The integration of Lean manufacturing, Six Sigma and sustainability: A literature review and future research directions for developing a specific model	180
3	Sustainable manufacturing-greening processes using specific Lean Production tools: An empirical observation from European motorcycle component manufacturers	157

Stage 3: Citation context analysis

Scite was first introduced in 2012 and is now a monitoring tool for both the positive and negative effects of the COVID-19 pandemic (Khamsi, 2020). Using Scite, artificial intelligence (AI) enabled tool that can help us evaluate evidence and correlation between the opinions of different authors.

Specifically, Scite is a platform for discovering and evaluating scientific papers via citation context analysis, providing citations based on context, showing whether the statement provides supporting or contrasting evidence for a referenced work, or simply mentioning an idea.

3. Result and Discussion

Lewis (2000) pointed out that sustainability should consider continuous innovation, the evidence suggests some companies that are considered to have achieved the strongest Lean production innovation performance do not prove superiority and do not reach success in terms of profit, on the other hand, the pressure surrounding innovation in production equipment creates unintentional pressure on the company annual's revenue.

Lewis researched multiple case studies to evaluate the effect of Lean manufacturing practices on an organization's competitiveness. The three companies at the center of the investigation, two of them supported the hypothesis that organizations applying Lean will result in an overall decline in innovative performance. This study provides empirical support for the suggestion that the more successful Lean thinking is applied to improve processes, the more focused the organization tends to be on incremental production changes, and the fewer innovation activities of firms are involved. Therefore, Lean is not qualified to provide a sustainable competitive advantage from Lewis's point of view.

Machine learning and artificial neural network tool reveal a few studies (2 papers) support research results of Lewis (2020) (Figure 3). Although Lean has been linked to core sustainable business development, there is a trade-off between Lean and sustainable; there is no evidence to support finding that when one of them is implemented, it will automatically lead to progression of the other. The benefits of Lean performance may decrease or increase depending on external conditions and factors, especially the influence of market power (Priyono and Idris, 2018). Besides, the study of Wu et al. (2018) pointed out, short-term Lean management practices improve long-term sustainability performance whistle long-term Lean management practices do not make sense for short-term sustainability performance, the more balanced sustainable practice is, the stronger and healthier we will be. However, we do not think their viewpoints provide persuasive evidence. Regarding sustainability theoretical framework background, Wu et al. (2018) mention "Organizational ambidexterity" which has been widely discussed in the context of sustainable competitive advantages (innovation, adaption, enterprise strategy), nevertheless when considering time dimension, "less attention has been focused on sustainability-related organizational ambidextrous activities".

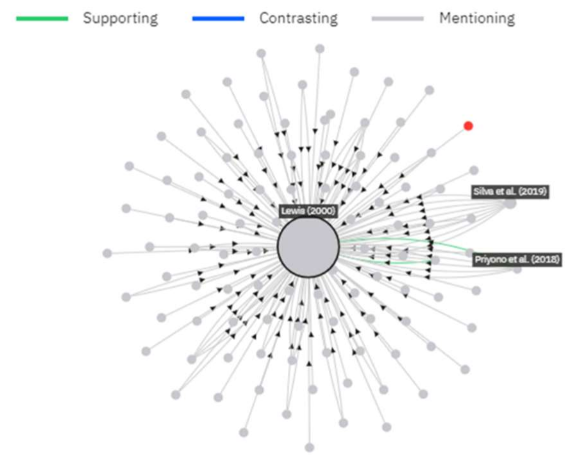


Figure 3. Context of citation

Although there are not many opinions to support Lewis's point of view, the figure shows that there are no contrary opinions about Lewis's research results; the remaining studies partially mention the opinion of the author Lewis. The result can be justified because the authors did not provide any evidence; they just repeated Lewis's conclusion or statement. Furthermore, the observation reinforces the view that almost all research concerns Lean towards environmental impacts.

4. Conclusion

The research contributed to the existing knowledge by exploring the sustainability paradigm. It is not persuasive enough to claim to Lean itself to gain sustainable competitive advantages. Based on research findings, we temporarily conclude that Lean is only a necessary condition, not a sufficient condition to achieve sustainability. The term Lean and sustainability (competitive advantages) should be separated. In the future, the need for empirical research should confirm the linkage of Lean and sustainable competitive advantages.

Reference

Abualfaraa, W., Salonitis, K., Al-Ashaab, A., & Ala'raj, M. (2020). Lean-Green Manufacturing Practices and Their Link with Sustainability: A Critical Review. *Sustainability*, *12*(3), 981.

Hartini, S., & Ciptomulyono, U. (2015). The Relationship between Lean and Sustainable

Manufacturing on Performance: Literature Review. *Procedia Manufacturing*, *4*(1), 38-45.

Henao, R., Sarache, W., & Gómez, I. (2019). Lean manufacturing and sustainable performance: Trends and future challenges. *Journal of CLeaner Production*, 208, 99-116.

Holweg, M. (2006). The genealogy of Lean production. *Journal of Operations Management*, *25*(2), 420-437.

Iranmanesh, M., Zailani, S., Hyun, S., Ali, M., & Kim, K. (2019). Impact of Lean Manufacturing Practices on Firms' Sustainable Performance: Lean Culture as a Moderator. *Sustainability*, *11*(4), 1112.

Kaufmann, G. (2020). Aligning Lean and value-based management (p. 16). Cham: Springer.

Khamsi, R. (2020). Coronavirus in context: Scite.ai tracks positive and negative citations for COVID-19 literature. Retrieved 16 April 2021, from https://www.nature.com/articles/d41586-020-01324-6.

Krajnc, D., & Glavič, P. (2005). A model for integrated assessment of sustainable development. Resources, *Conservation and Recycling*, *43*(2), 189–208.

Lewis, M. (2000). Lean production and sustainable competitive advantage. *International Journal of Operations & Production Management*, *20*(8), 959-978.

Liyanage, J. (2007). Operations and maintenance performance in production and manufacturing assets. *Journal of Manufacturing Technology Management*, *18*(3), 304-314.

Piercy, N., & Rich, N. (2015). The relationship between Lean operations and sustainable operations. *International Journal of Operations & Production Management*, *35*(2), 282-315.

Pirraglia, A., Saloni, D., & Dyk, H. (2009). Status of Lean manufacturing implementation on secondary wood industries including residential, cabinet, millwork, and panel markets. *Bioresrouces*, *4*(4), 1341-1358.

Priyono, A., & Idris, F. (2018). Analysing the adoption of Lean production in remanufacturing industry. *Journal of Industrial Engineering and Management*, *11*(4), 697.

Resta, B., Dotti, S., Gaiardelli, P., & Boffelli, A. (2016). Lean Manufacturing and Sustainability: An Integrated View. *IFIP Advances in Information and Communication Technology*, 659-666.

Shingo, S. (1996). O sistema Toyota de produção: o ponto de partida da engenharia de produção. Porto Alegre: Bookman.

Smith, Linda C. (1981). Citation context analysis. *Library Trends, 30*(1):83-105.

Vinodh, S., Sarangan, S., & Chandra Vinoth, S. (2014). Application of fuzzy compromise solution method for fit concept selection. *Applied Mathematical Modelling*, *38*(3), 1052-1063.

Wu, L., Subramanian, N., Gunasekaran, A., Abdulrahman, M., Pawar, K., & Doran, D. (2018). A two-dimensional, two-level framework for achieving corporate sustainable development: Assessing the return on sustainability initiatives. *Business Strategy and the Environment*, *27*(8), 1117-1130.

APPLICATION OF FAULT DETECTION METHOD AND PREVENTIVE ACTION TO IMPROVE SOLAR CELL PRODUCTION STAGE PERFORMANCE

Pham Ngoc Phung[1], Ho Thi Linh[2], Pham Anh Trong[3], Do Ngoc Hien[4]

[1] *Department of Industrial Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, pnphung.sdh19@hcmut.edu.vn*

[2] *Department of Industrial Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, htlinh.sdh19@hcmut.edu.vn*

[3] *Department of Industrial Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, patrong.sdh19@hcmut.edu.vn*

[4] *Department of Industrial Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, hienise97@hcmut.edu.vn*

Abstract. Saving cost is always one of the most important purpose in manufacturing. This study was executed at NPR (Negative Photo Resist) coating process in the solar panel factory in order to reduce downtime, failure rates and emergency repair costs. It is a very important step in the manufacturing process, which supports to isolate the layers in the panel and create more battery cells (the number of battery cells is proportional to the power of the battery). The failure rate of the broken panel in the NPR coating process results in damage to the coating roller causing stop the production line and affects the production and costs of replacing the coating roller. After applying fault detection method, designing support tools to replace coating roller, the research results show that the downtime, failure rates and emergency repair costs are significantly reduced.

Keywords: Work design; effective; solar panel manufacturing; operational efficiency

1. Introduction

A solar panel, or photo-voltaic (PV) module, is an assembly of photo-voltaic cells mounted in a framework for installation. Solar panels use sunlight as a source of energy and generate direct current electricity. A collection of PV modules is called a PV Panel, and a system of Panels is an Array. Arrays of a photovoltaic system supply solar electricity to electrical equipment.

At present most solar modules are produced from crystalline silicon (c-Si) solar cells made of multi-crystalline and monocrystalline silicon, while the rest of the overall market is made up of thin-film technologies using cadmium telluride (CdTe), CIGS and amorphous silicon (α-Si). Thin-film technology produce a relatively high-efficiency conversion for the low cost compared to other solar technologies. With their minimum material usage and rising efficiencies, thin-film technology is becoming more popular.

Thin-film photovoltaic cells was produced by coating thin material layers on the top of substrate glass. Then they cut these layers into string of solar cell by using laser. A solar cell can be considered as battery, so the efficiency of solar panels depends on the number of solar cells. Therefore, isolating layers stage is extremely important to make sure we can produce the solar panel with the maximum and most efficient solar cells possible.

Downtime due to machinery failure can happened at any production stage. This study applies the work design principle to determine which factor causing the most downtime on NPR coating stage to isolating thin-fim layer in a solar panel manufactory, thereby proposing methods to reduce downtime, failure rates and emergency repair costs. The layout of this article consists of five main parts: Introduction, Literature review, Research Methodology, Results and Conclusion.

2. Literature review

Nowadays, many companies are competing to produce the best products for consumer use. Instead of having customer needs as a priority they also compete in providing the highest quality products. Most companies are now able to produce a large quantity of product at a much lower price. However, this leads to trade-offs in product quality. In keeping with the quality of products, companies should eliminate products rejects due to faulty, "Poka Yoke" is a Japanese term that means "mistake-proofing". It is a method used to ensure the quality of a product can be controlled by preventing errors. It will not only prevent defects from occurring, but it also can expect errors and techniques to control it

(Dudek-Burlikowska & Szewieczek, 2009). Offering fail-safe solutions, Poka Yoke allows relevant processes to run smoothly (Shingo, 1987), (Velmanirajan et al. 2013)

Isaksen (1998) note that there are various methods used to find solutions to problems such as mind mapping, brainstorming, lateral thinking, etc. These methods are used to identify problems and the root cause, but they are lacking on the capability to literally show the solutions to the problems. On contrary, TRIZ methodology assists US not only to identify the problems but also provides direct solutions to the problem along with confidence that possible new solutions to the problem have been considered (Ilevbare, Probert, & Phaal, 2013).

Helmi et al. (2017) using the concept of Poka Yoke, which prevent employees from doing errors while avoiding repetition work can reduce their burden and at the same time smoothen the carbon fibre strips production. The use of paper cutter design knife to replace the scissors has given a major impact as in Poka Yoke applications for this cutting process.

The errors can be committed by a human operator in the control program. When a command error is made, the presence of a product in the system can cause damage. To prevent human control errors, a robust filter be designed to place inside the PLC (Programmable Logic Controller) which authorizes or forbids outputs from the PLC. The filter is composed of several logical constraints which have to be respected at each PLC cycle (Manrangé et al. 2010). Besides, supervisory circuits are designed for various type of application. An Integrated Microprocessor Supervisory Chip was specifically designed for Microprocessor peripheral devices for Monitoring Power Failure (Intan, 2010).

Wang et al. (2010) confirm that in manufacturing, maintenance task becomes more and more dependent on supporting equipment. When supporting equipment breaks down, maintenance task can not keep on proceeding, maintenance delay time gets prolonged. This phenomenon was much more common especial on some sophisticated and high-price support equipment. In order to solve the problem, more and more institutions have accentuated the impact of support equipment failure bringing on logistics process. The method of analyzing impact support equipment bringing about was elaborated and got proofed through an application.

In the work area, workload demand is the factor that most influences employees' work stress (Sprigg & Jackson, 2006). In addition, to minimize the impact of musculoskeletal disease on employees, workplace design needs to consult with anthropometric work and allow them to rotate from stress to less stressful work at work a certain time (Estember & Aguila, 2020). Therefore, in maintenance activities, when applying support tools, it is very important to evaluate the anthropometry to suit the staff's operations.

3. Research Methodology

First, to detect problems or issue in the workplace, the Gemba walk method was used to approach the problem. When a problem occurs, it is necessary to go to the field to observe, understand and collect information as accurately as possible instead of sitting on the office and request indirect information. Gemba walk is a method that helps managers go directly to the work area to observe, analyze data to identify improvement opportunities, and get quick feedback from employees. Gemba Walk's goal is to get quick and easy feedback, identify concerns related to the workplace or process, build trust with the team by listening their problem, understand the value stream and its problems rather than looking at the results or making superficial judgments.
In the workplace area, in the last 6 months of 2020, the production lines faced with many major downtimes. However, analyzing the downtime is not easy. Therefore, it is extremely important to collect accurate information about the problem related to the production line stoppage.

After classifying what is an unplanned downtime, what is a planned downtime, the manager needs to calculate the downtime according to each fault category to simplify the collected data. In general, we choose to only drive into unplanned downtime to focus on resolving issues. The goal of this study was to identify the cause and reduce downtime on the production line.

During the analysis, this problem usually consists of two main types: Local failure resulting in the machine must be stopped for maintenance or minor disruption to the operation process.

Local failure is the most visible loss in production because of its irregularity and its obvious impact on continuous production in the factory. Cases of this loss include mold/die failures, unplanned maintenance operations, general mechanical/electrical failures or parts of equipment or non-operating equipment cases as required (function/technology parameters). Pause stops, to distinguish them from local malfunctions, often include incidents that cause short downtime – for example, less than 5 minutes – and often do not require the presence of a technician. Technical/maintenance.

Actual situations may include problems with on-line flow, jamming/clogging, problems with feeding/guiding parts, wrong supplies/materials, obscured sensors, incidents small in the following stages, etc (Shagluf et al. 2012.

After identifying the local failure that caused the biggest downtime and its cost impact, we performed Root cause analysis (RCA) to find the cause of the equipment failure on this line.

RCA is a very effective method used to find the root cause of incidents, errors or an unexpected result that has occurred. This method focuses on identifying system and process failures, not individual errors.

The main purpose of RCA is to identify the root cause of an incident, error, or accident, in order to take effective corrective action, prevent it from happening again, or successfully resolve the problem ("Success" means almost certain to prevent the problem from happening again). Incident analysis also helps the organization to detect risk factors and weaknesses in operational processes for timely remedial action (Ammerman, 1998). RCA also plays a key role in preventing particularly serious incidents as sometimes we make improvements but are not effective because the root cause is not changed.

To identify all the causes involved in the entire process of equipment failure on the line, we apply the following tools:

- 5 WHY (Identify Contributing Process Factors) continuously answers the question why, according to studies, if the 5 WHY model is implemented, the answer will be found.
- In addition, we must collect downtime/cost data to expose the cause of the failure.

After analyzing the main causes of downtime on the production line, we apply the Poka Yoke method to reduce the failure rate on the equipment.

Poka Yoke is the Japanese term, meaning error prevention or failure prevention, developed by management consultant Shigeo Shingo as part of the Toyota production system. The purpose of Poka Yoke is to design a process to help detect errors, prevent and correct them immediately, eliminate defects at the source or promptly warn when they occur. Poka Yoke is made using simple objects like furniture, jigs, gadgets, warning devices to prevent people from making mistakes. Stopping equipment for planned maintenance is part of the job of a manufacturing business. However, the equipment should not be stopped urgently, the equipment will only stop when we actively stop it (Robinson, 1997).

The sequence of work to be performed when applying error protection includes the following steps:

Step 1: Identify errors that can occur even with preventive actions. Review each step in the process and ask, "What is the most likely error in this step, human error or equipment error?"

Step 2: Decide on a method to detect some possible or imminent mechanical failure or failure.

Step 3: Identify and select appropriate corrective action when errors are detected:

- Control: activities aimed at self-correcting process errors.
- System Stop: The device stops or terminates the process when an error occurs.
- Warning: alerting people involved in the work when an error occurs.

Error prevention tool (Poka - Yoke) is used to eliminate problems and errors in the process, helping the process to maintain stability. Error prevention tools have 3 main functions: Detect, Fix and Prevent. (Shingo S. , 1988)

In this study, when an error occurs, the Poka - Yoke tool will automatically disconnect the working system and alert the relevant people. Timely detection of errors and incidents helps to prevent defective products, and at the same time minimizes the cost of damage and repair of equipment on the production line. Poka Yoke's role in this study is: to help reduce production costs, through reducing the cost of replacing damaged components; reduce management costs, no labor required for monitoring or error handling, saving time for activities that bring higher value.

In addition, after detecting defective products, it is also difficult to quickly replace these products off the line, as they are heavy objects. From an anthropometric perspective, allowing workers to move these objects will result in cumulative back and shoulder injuries. Therefore, we have proposed to use support tools, both to ensure the long-term health of workers, and to meet the needs of quickly replacing defective products. The application of modern machinery and equipment to assist in error handling and maintenance will help businesses save time, the number of workers, ensure the health of employees, and this also collectively known as improving production efficiency.

4. Results

4.1. Current Status:

NPR coating process in the manufactoring which was chosen to conducted improvement has layout as below:

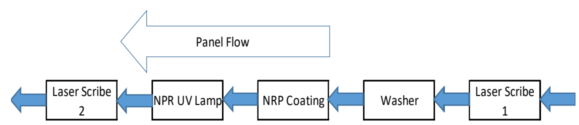


Figure 1. NPR coating process

According to above process, the solar panel wil be separated into smaller solar cells by Laser Scribes 1, 2. The purpose of this Laser Scribe process is to reduce current produced by the panel and increse voltage to meet almost inverter requirements. And we also need to isolate Laser Scribe 1 and Laser Scribe 2 to make sure there is no short circuit which will cause power loss. The NPR (Negative Photo Resist) Coating process will help us on this. There is a roller used a reverse/reverse application method to apply a target of 1µm of NPR liquid to the entirety surface of the panel. Material using at this process crosslinking into a very hard substance similar to a ceramic when exposed to ultraviolet light. This reaction known as polymerization, change from soluble to insoluble. NPR UV tool will help us on this stage. Once the panel exits the NPR UV, it is ready for the next process.

Downtime at each production stage in the last 6 months of 2020 is shown as below:

Table 1. Downtime and Emergency repair cost at production stage in the last 6 months of 2020

Tool	Downtime (hrs)	Estimate Emergency Repair Cost (USD)
Washer	47.65	5,390
NPR coating	**65.49**	**41,350**
NPR UV	5.65	95
Laser	10.26	36,900

We can find out that almost downtime causing stop the production line happened at *NPR Coating* process. Taking a deep drive into the total downtime of *NPR Coating*, analysising their falure mode, we got the chart as below:

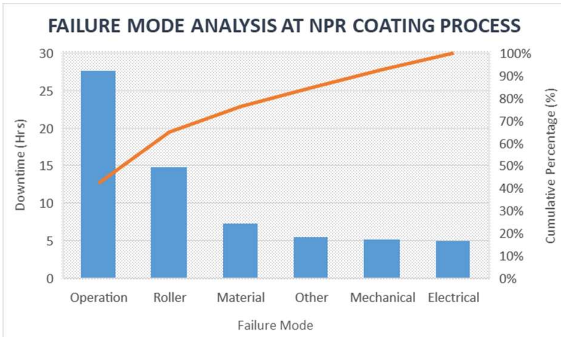


Figure 2. Failure mode analysis at NPR Coating

We cannot reduce downtime for *Operation* activities because we have to stop the production line and change the material to keep the machine running. It is compulsory downtime so we will focus on issues related to roller instead.

All the times the problem occurs in the roller because the broken glass scratches the roller surface and as a result, we have to replace the roller to not affect the quality of the panel. Broken, chipped, or skew panel when entering the roller are all at risk of scratching the roller surface. In the last 6 months of 2020, there were 4 roller replacements due to broken panel with an average downtime of **3,705 hrs** each and around **7,300 USD** cost for each new roller.

Table 2. Downtime due to broken panel in 2020

No	Month	Downtime (hrs)
1	June	4.34
2	July	4.62
3	October	4.52
4	December	1.34

4.2. Improvement proposal:

To minimize the frequency of roller replacement, we must deal with the root cause that causing broken roller: the broken glass inside the roller.

There is two type of broken glass causing broken roller:

1) Panel which is already chipped or broken before entering roller.

2) Good panel but it becomes skew while being transferred on conveyor. When it enters rollor, it breaks under roller pressure while it is working.

Prevent chipped or broken panel entering NPR Coating provess

We planned to install camera with computer computer integrated image processor. This tool was intalled before the washer in the direction of panel flow. We chose this position but before NPR coating because almost chipped or broken panel come from Laser Scribe 1 pre-process. And there was record that sometime broken panel stuck inside washer and we have to stop the production for cleanning broken panel. Another reason for this chose is restriction of space between NRP Coating and Washer machine, that is being used for another improvement (4.2.2).

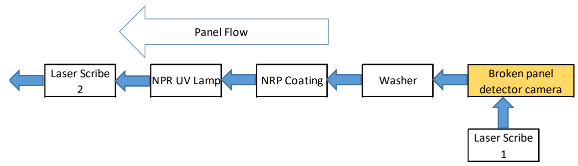


Figure 3. Install camera with computer computer integrated image processor

With this tool we can early detected chipped or broken panel and send the control signal to entry conveyor of the Washer. This tool use 2 cameras to capture image of panel in real time then compare with reference image. If there is abnormal, chipped or broken panel, processor will send signal to controller. When the signal is triggered, conveyor will stop and ring an alarm so that nearby operator can have quick action to remove broken panel then reset alarm and return the line to production.

Figure 4. Cameras before the washer in the direction of panel flow

Prevent skew panel:

To prevent skew panel entering roller coat, we designed an array of sensors that detects edge of the panel. We placed those sensor at entry conveyor of NPR coating like a final gate keeper who prevent not good panel from entering roller.

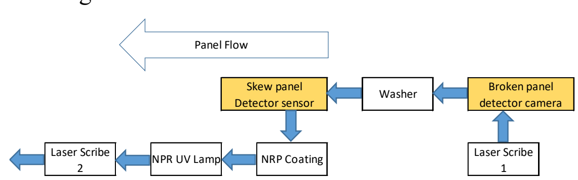


Figure 5. An array of sensors detecting edge of panel

There are 4 sensors fixed above the conveyor, parallel to the edge of the panel in the direction of panel flow. When the well aligned panel come, all 4 sensors will capture the signal at the same time. If not, the signal of these 4 sensors will be on uneven, based on that we can detect the skew glass.

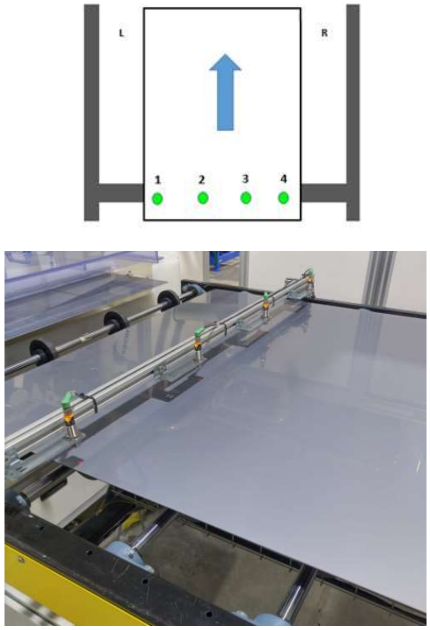


Figure 6. Four sensors capturing the signal

Reduce downtime when replacing roller:

In the past, every time we replace the roller, we have to use forklift to lift the roller. This take us a lot of time to maneuver the forklift. However, the forklift is also difficult to manipulate in narrow spaces in the floor. In the last 6 months of 2020, there have been 3 times of replacing rollers with this method with an average time of 4.49hrs.

Figure 7. Replace the roller using forklift

In December 2020, after observing and assessing the difficult situation when replacing those rollor, we had a improvement proposal that using crane instead of forklift.

Figure 8. Replace the roller using crane

We used this method for roller replacement in December 2020. This time we spent only 1.34 hrs compared to 4.49 hrs on average of previous times.

4.3. Results after performing improvement:

Summary of implementation date and cost for each method is as follows.

Table 3. Summary of implementation date and cost for each method

Method	Implementation Date	Cost (USD)
Broken Panel Detector Camera	Jan 2021	~20,000
Skew Panel Detector Sensor	Jan 2021	1,735
Using crane	Dec 2020	2,050
Total		**23,785**

Broken Panel Detector Camera and Skew Panel Detector Sensor have been applied in January 2021. Up to now, they detected 4 times of chipped or broken glass; 3 times skew glass and thanks to that we have prevented the risk of roller damage before it happens. Roller has never been damaged in this year.

Compared to the defect rate in the second half of 2020, with repair costs of about $41,350 - we have saved about $17,500 in the first half of 2021. And will continue to save about $82,000 per year if the improvement method continues to prove its effectiveness.

5. Conclusion

Work design is a set of function of industrial and organizational psychology that focus on improving productivity, effectiveness; reducing cost of a process. Based on Work Design theory, we have implemented some improvement on a stage of the process in the solar panel factory. The results obtained are very positive:

- Early detect almost issue and prevent them from happening. Thus, the failure rate is significantly reduced and there is no need for emergency repair costs.

- Reduce around 70% time for repairing machine when it issue occurs.

- No manufacturing process is perfect, but if we keep chasing continuous improvement we can achieve excellence.

Acknowledgments

We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study

References

Ammerman, M. (1998). *The Root Cause Analysis Handbook: A Simplified Approach to Identifying, Correcting, and Reporting Workplace Errors.* Productivity Press.

Dudek-Burlikowska, M., & Szewieczek, D. (2009). The Poka-Yoke method as an improving quality tool of operations in the process. *Journal of Achievements in Materials and Manufacturing Engineering*, 95-102.

Estember, R. D., & Aguila, D. K. (2020). The Effects of the Workplace Design, Work Methods Design and Health Problems and Occupational Diseases in the Prevalence of Musculoskeletal Disorders among Bank Tellers. *IEEE 7th International Conference on Industrial Engineering and Applications.*

Helmi, S. A., Nordin, N. N., & Hisjam, M. (2017). Syed Ahmad Helmi, Nur Nashwa Nordin, and Muhammad Hisjam. *3rd International Materials, Industrial and Manufacturing Engineering Conference.*

Ilevbare, I., Probert, D., & Phaal, R. (2013). A review of TRIZ, and its benefits and challenges in practice. *Technovation*.

Intan, N., Amir Abas, M., & and Sakrani, S. (2010). Design an Integrated Microprocessor Supervisory Chip for monitoring power failure. *34th IEEE/CPMT International Electronic Manufacturing Technology Symposium.*

Isaksen, S. (1998). *A Review of Brainstorming Research.* New York: Creative Research Unit, Creative Problem Solving Group-Buffalo.

Marangé, P., Benlorhfar, R., Gellot, F., & Riera, B. (2010). Prevention of human control errors by robust filter for manufacturing system. *IFAC Proceedings Volumes*, (pp. 175 - 180).

Robinson, H. (1997). Using Poka-Yoke Techniques for Early Defect Detection. *Sixth International Conference on Software Testing Analysis and Review.*

Shagluf, A., Longstaff, A. P., & Fletcher, S. (2012). *Effective Maintenance Management - Risk and Reliability Strategies for Optimizing Performance.* Industrial Press.

Shingo. (1987). *The Sayings of Shigeo Shingo: Key Strategies for Plant Improvement (Japanese Manage).* The Nikkan Kogyo Shimbun Ltd.

Shingo, S. (1988). *Zero Quality Control: Source Inspection and The Poka-yoke System.* Productivity Press.

Sprigg, C. A., & Jackson, P. R. (2006). all Centers as Lean Service Environments: Job-Related Strain and the Mediating Role of Work Design. *Journal of Occupational Health Psychology*.

Velmanirajan, K., Rajaraman, G., Karthikeyan, K., Dinesh, D., & Leyland, A. (2013). Lean Manufacturing in Chassis Assembly through Poka-Yoke. *International Journal of Technology Enhancements and Emerging Engineering Research (IJTEEE)*.

Wang, Y., Wang, N., Liu, R., & Kang, R. (2010). An Method for Maintenance Waiting Time Considering Support Equipments Failure. *International Conference on System Science, Engineering Design and Manufacturing Informatization.*

AN OPTIMIZATION MODEL FOR VEHICLE ROUTING PROBLEM WITH MULTIPLE PICKUP AND MULTIPLE DELIVERY WITH HETEROGENEOUS VEHICLE AND TIME WINDOW

Phan Nguyen Ky Phuc[1] and Nguyen Le Phuong Thao[2]

[1] Department of Industrial and Systems Engineering at International University, Ho Chi Minh City, Vietnam, pnkphuc@hcmiu.edu.vn

[2] Department of Industrial and Systems Engineering at International University, Ho Chi Minh City, Vietnam, nlpthao@hcmiu.edu.vn

Abstract: With the rapid development on E-logistics industry, there are several transportation models modified to meet the customers' requirement on the aspect of speed and coverage which are stricter due to the highly competitive logistics market. One of the widely used logistics network to tranship the goods among regions is which requires that each vehicle must visit some pickup nodes first, for instance, warehouses to pick up the orders then make deliveries for customers in the list. Each pickup node has its own list of more than one requiring delivery. The objective is to minimize the total travelling cost while real-case constraints are still considered. This study focuses on solving the mentioned vehicle routing problem (VRP) of E-logistics service providers and proposes a mathematical model for this multiple pickup and multiple delivery vehicle routing problem with time window and heterogeneous fleets (NPMDVRPTWHF). This proposed optimization model is quietly flexible so it can be applied to not only E-logistics problems but also reserved logistics, retail distribution and school bus routing problems as well.

Keywords: Vehicle Routing Problem with time-window, heterogeneous fleets, multiple pickups, multiple deliveries, E-logistics

1. Introduction

As defined by Beamon B. (1998), supply chain is a manufacturing process in which raw materials are taken from suppliers and transformed into finished goods to deliver to final customers. Supply chain management plays an extremely critical role in any business operations. It guarantees that the finished products can be delivered properly to the customers. And logistics is one of the determinants on the total operation cost. The Economic and Social Commission for Asia and the Pacific (ESCAP) has proposed an estimate method to yield a statistical solution that the most occupied cost in logistics cost is transportation cost which takes up more than 60% of the total cost in consisting of costs for shipping the items from the factories and retailers in the distribution network to the final customers. Therefore, it is vital for businesses to consider more about transportation cost optimization to provide a competitive price while some real-world application constraints are still considered such as required service time interval, a resource of heterogeneous fleets of vehicle, and some unpredicted events, etc. Additionally, with the rapid development of sciences and technologies, E-logistics and last-mile delivery has emerged to meet the higher requirements of customers in term of goods transportation.

Therefore, vehicle routing problem (VRP) began to be studied by many researchers. It is a problem finding a set of routes with the objective of operation cost minimization, which can provide businesses solution to solve headache logistics cost optimization. This study aims to study about a variant of vehicle routing problem, Vehicle routing problem with multiple pickup and multiple delivery and with a set of constraints including time window and heterogeneous fleets of vehicle usage, which can be applied in E-logistics field and others with suitable model networks. The objective of the study is to build a proper mathematical model for solving such called type of problem.

2. Literature Review

The vehicle routing problem concerns the problem seeking the optimal set of routes with minimum cost or travelling time, given a set of nodes as customers and one or more depot where the vehicle departs there and set of associated constraints. It was first introduced by G.B Dantzig and J.H. Ramser. (1991) in the truck dispatching problem. Their proposed problem was inspired by a gasoline transportation problem in which the objective was to find the shortest routes to the gas station to deliver gasoline, they published the first mathematical formulation and proper approach to the problem. A few years then, an effective greedy heuristics method was introduced by Clarke and Wright improving Dantzig-Ramser approach. Then, diversified research based on the two mentioned papers were proposed to solve VRP problems including approximate and exact approaches.

Reviewed by J.R. Montoya – Torres et al. (2015), several variants of VRP with different practical constraints consideration were studied and published. Capacitate VRP is one of the most common versions of VRP in which the vehicles are homogeneous, there are a single depot and a set of delivery requesters associated with demand quantity which cannot be split. The featured constraint of this problem is vehicle capacity constraint which the total carrying on the vehicle cannot exceed its capacity. The real cases then required to have multiple depots and time-window for each customer, the model of multiple depot vehicle routing problem with time-window was studied. The VRP with time-window (VRPTW) still considers a set of delivery requesters but adding time data related to a service time interval $[a_i, b_i]$ in which requester i is available to be served, serving time including loading and unloading time s_i, for the case the vehicle must wait to serve, waiting time a_i is included. With the development of supply chain network, another version of CVRP is the VRP with Backhauls (VRPB) was introduced to adapt the case which requires the vehicle revisit the customer nodes to pick-up items and travel back to the depot. In the problem, there are wo subset of customers. The first subset is the linehaul set containing n customers demanding a given quantity of goods to be delivered. And the second one is the backhaul set containing m customers having items required to be picked up and taken back to the depot. The new requirement in this problem is that the linehaul customer set must be served before the backhaul one. One more common variant of VRP is VRP with Pickup and Delivery (VRPPD). This variant is another version of CVRP, the new point is that each requester i has its pickup point P_i and an associated delivered quantity p_i picked up from P_i, and a destination point D_i where a delivered quantity q_i picked up from customer i and delivered to D_i

For small-sized problems, VRP can solve them effectively adopting exact solution methods introduced by Toth and Vigo (2002) and Baldacci *et al.* (2010). Besides, for those problem having large-size node set, researchers have introduced many heuristics and metaheuristics approach to obtain near-optimal solution. Some heuristics approach such as the space-filling Curve with optimal portioning (SFC-OP) heuristics, location-based heuristics (LBH), three-phase heuristics developed by combining a heuristics-based clustering algorithm are introduced to solve VRP. Several important and state-of-art modern heuristics are reviewed in the study by Cordeau *et al.* (2002) and Szeto *et al.* (2011).

3. Mathematical model

3.1. Problem identification

This study aims to build a mathematical model for such a new variant of VRP which is vehicle routing problem with multiple pickup and multiple delivery with heterogeneous fleets and time-window (MPMDVRPTWHF). The model is constructed by giving a directed graph including customer set as vertices set, $C = \{0, \dots, n + m + 1\}$ where node 0 represents the departure depot and node $(n + m + 1)$ corresponds to the ending depot of any route. The remaining nodes in the customer set are divided into 2 subsets: Pickup set P with n customers requiring pickup service and Delivery set D with m customers requiring delivery service. In the studied topic, to meet the constraint of classical VRP that each node is served exactly once, if the node requires both delivery and pickup, that node will be duplicated, for the first might be for the pickup node and another is delivery one. Therefore, in the set P and D, there could be a set of real customers and a set of dummies. The resource involves a set of heterogenous vehicle fleets K with M vehicles. Each vehicle is associated with an operation cost z_k and capacity Q_k. Each customer i may require a quantity of goods to be served q_i which is larger than 0 if it is pickup quantity and less than 0 if it is delivery quantity; required time interval to be served called time window interval (e_i, l_i) and service time s_i for loading and/or unloading process. Each arc travelling from node i to node j is associated with travelling time t_{ij} and distance d_{ij}. Each route must not exceed the total travelling time t_{max}. There is a relation set of nodes showing each pickup node has the associated set of delivery nodes, and the parameter R_{ij} show that relationship with binary numbers. In the problem, x_{ij}^k will be the decision variable on which trail will be chosen by vehicle k and on the optimal solution. The other two variables are T_i^k (in hour) and L_i^k (in kilograms) representing the start-time of service and current load at node i of vehicle k.

The application of this type of model is not restricted in E-logistics network, where at each warehouse in a city, it always has products which are collected from the sellers need to be picked up and delivered to other warehouses having customer nearby, while having a quantity of items received from others; but it can also apply to any VRP cases having network characteristics that a node requires a quantity of products to be delivered and a quantity of products to be picked up, such as retailer network routing problem, in which , bus scheduling problem where at each bus station, there are a set of students needing to be picked to their schools and a set of students needing to get off the bus; and reversed logistics as well where at each factory, while demanding to receive materials, it can also have empty materials containers that need to be collected to be recycled at the depot.

3.2. Mathematical formulation

Sets:

C The customer set, $C = \{0, \ldots, n+m+1\}$

C_1 The customer set not including the starting depot, $C_1 = \{1, \ldots, n+m+1\}$

C_2 The customer set not including the ending depot, $C_2 = \{0, \ldots, n+m\}$

K The vehicle set $K = 1, .. M$

P Pickup set, $P = \{1, .. n\}$

D: Delivery set $D = \{1, \ldots, m\}$

R: Relationship set (i, j) in which i is the pickup point of j

Parameters:

qi Pickup or delivery quantity of customer i

s_i The service time at the customer i

e_i The earliest start time at customer i

l_i The latest start time at customer i

d_{ij} The distance between customer i and j

Q_k The capacity of vehicle k

z_k Cost per distance of vehicle k

t_{max} The maximum working duration

M The number of vehicles

t_{ij} The traveling time from node i to j

R_{ij} Binary variable, $R_{ij} = 1$ if node i is a pickup point and node j is a delivery point.

$BigM$ The large constant number

Variables:

x_{ij}^k The 0-1 decision variable, if vehicle k travels directly from node i to node j, then $x_{ij}^k = 1$, otherwise, $x_{ij}^k = 0$

T_i^k Start time of service of vehicle k at customer i

L_i^k Load of vehicle k at customer i

Objective function: (1)

$$\sum_{k \in K} \sum_{i,j \in C} x_{ij}^k d_{ij} z_k$$

Subject to: (2)

$$\sum_{k \in K} \sum_{j \in C_2} x_{ij}^k = 1,$$

$$\forall i \in C \backslash \{0, m+n+1\}$$

$$\sum_{h \in C_2} x_{hi}^k - \sum_{m \in C_2} x_{mj}^k = 0, \quad \forall\, k \in K, \forall\, (i,j) \in R \quad (3)$$

$$\sum_{j \in C_1} x_{0j}^k \leq 1, \forall k \in K \quad (4)$$

$$\sum_{j \in C_2} x_{j,n+m+1}^k \leq 1, \forall k \in K \quad (5)$$

$$\sum_{j \in C_2} x_{ji}^k = \sum_{h \in C_1} x_{ih}^k, \quad (6)$$

$$\forall i \in C \backslash \{0, n+m+1\}, h \neq i, j \neq i, \text{k} \in K$$

$$T_j^k \geq T_i^k + (t_{ij} + s_i) - BigM(1 - x_{ij}^k), \quad (7)$$

$$\forall k \in K, i \in C_2, j \in C_1$$

$$L_j^k \geq L_i^k + q_j - BigM(1 - x_{ij}^k), \quad (8)$$

$$\forall i \in C_2, j \in C_1 : i \neq j, k \in K$$

$$T_i^k + R_{ij}(t_{ij} + s_i) \leq T_j^k, \quad (9)$$

$$\forall k \in K, i \in P, j \in D$$

$$e_i \leq T_i^k \leq l_i, \quad (10)$$

$$\forall i \in C \backslash \{0, m+n+1\}, k \in K$$

$$\max(0, q_i) \leq L_i^k \leq \min(Q_k, Q_k + q_i), \quad (11)$$

$$\forall i \in C, k \in K$$

$$T_{n+m+1,k} - T_{0k} \leq t_{max}, \forall k \in K \quad (12)$$

$$\sum_{k \in K} \sum_{j \in C} x_{0j}^k \leq M \quad (13)$$

$$x_{kij} \in \{0, 1\}, \forall\, i, j \in V, k \in K$$

In the formulation, equation (1) shows the objective function of the model which is operation cost minimization. The constraint that each node must be served at least by one vehicle is guaranteed in equation (2). The requirement that if there is an order between pickup node i and delivery node j, this order must be served by the same vehicle k is ensured by equation (3). Equation (4) and (5) claim that any used truck must visit the source depot and destination depot at least once. The constraint that each vehicle visits a node, then it also must leave that node is claimed in the equation (6). Set of Subtour elimination, time constraint and load constraints are integrated in the equation (7) and (8). The requirement that each pickup node must be served before its set of delivered requesters is shown by equation (9). Time-window and capacity bound constraints are presented in equation (11). The last two constraints (12) and (13) limit the driver working duration and the number of vehicles used.

4. Numerical experiments

4.1. Benchmark description

The instance set adopted to validate the studied model is published by Goeke (2017) which is purposed to solve Pickup and delivery problem with time windows and electric vehicles (PDPTW-EV) and is also applicable for the vehicle routing problem with time window (VRPTW) and green vehicle routing problem (G-VRP). The benchmark instance has two datasets: a set of 36 small-sized customer sets which is ranged from 3-9 customers and a set of 56 larger-sized set with 50 requesters. Due to the limitation of the CPLEX software on the problem size, the small-sized sets will be considered only to validate the mathematical model. There are still some necessary modifications on the given instances to fit the proper data form to be input to the model. Specifically, there are some more data needed to be added to the instances which are related to heterogeneous fleets, pickup-delivery relationship and working duration. The added vehicle parameters include its identifications, corresponding capacity, and unit cost. hour per trip to obey the labour policy, this type of data is used in the working duration constraint.

4.2. Computational result

This section aims to validate the proposed mathematical model. To evaluate the validation of the solution, CPLEX solver and the benchmark detailed in section 4.1. will be used to test the model on Intel® Core (TM) i5-8265U CPU @ 1.60GHz 1.80 GHz. Computational time on solving the problem on CPLEX solver is averagely around 23.78s when solving the problem size from 6 to 18 customers. This mathematical model can solve the small-sized problems efficiently both in computational time and solution optimality.

Table 4.1 Data set for customer set.

Point	Service time	Pickup/ Delivery demand	Earliest time	Latest time
0	0,25	0	0	100
1	0,25	56	7	9
2	0,25	15	7	9
3	0,25	19	7	9
4	0,25	-25	7	9
5	0,25	-31	7	9
6	0,25	-7	9	13
7	0,25	-8	9	13
8	0,25	-9	9	13
9	0,25	-10	9	13
10	0,25	0	9	13

Table 4.2. Data set for vehicle set.

Vehicle No.	Vehicle capacity	Cost per mile
1	50	5500
2	50	5500
3	100	7000
4	100	7000
5	100	7000
6	100	7000

Table 4.3. Data set for relationship between pickup points and their delivery points

Pickup	Delivery
1	4
1	5
2	6
2	7
3	8
3	9

To illustrate the data requirements and how the output of the model will be displayed, a generated data set including 1 depot, 3 pickup points, and 9 delivery points. The associated parameters related to carrying quantity, service time and time-window intervals are demonstrated in the table 4.1. The vehicle parameters related to maximum capacity and cost per mile for the heterogeneous fleets of vehicle are displayed in table 4.2. Table 4.3 gives the set of delivery requesters associated with each pickup node. There are also a distance table and travelling-time table for every pair of nodes not shown in this paper. The optimal routing result is shown graphically in the figure 4.1.

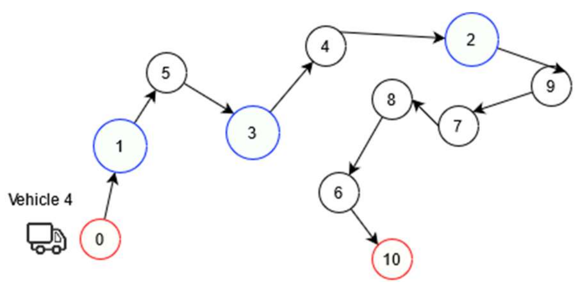


Figure 4.1. Graphical result for the 11-customer data set

5. Conclusion

A variant of vehicle routing problem with time window was studied in this paper, which considered a network of nodes constituted by multiple pickup and multiple delivery nodes and utilized a resource of heterogeneous fleets of vehicles, and working duration allowed. Those added constraints were formulated to increase the applicable features for the study. Therefore, the proposed mathematical model is applicable for those networks having multiple pickup and multiple delivery requesters, or even the case having customers requiring multiple visits such as E-logistics routing problem, reversed logistics routing problem, retail network routing problem, etc. The model is validated by a famous benchmark instance used for VRP and run by CPLEX software. Therefore, the proposed mathematical model is reliable to be applied in such small-sized cases with the proper transportation network.

For larger-sized cases that CPLEX software becomes less efficient in term of computational time, heuristics or metaheuristics approach should be studied and combined with the proposed mathematical model to obtain approximate solutions within an acceptable tolerance value.

Reference

Anh, P.T.; Cuong, P.T.; Ky Phuc, P.N. (2019). *The vehicle routing problem with time windows: A case study of fresh food distribution center.* In Proceedings of the 11th International Conference on Knowledge and Systems Engineering (KSE), Danang, Vietnam, 24–26.

Baldacci, R.; Toth, P.; Vigo, D. (2010). *Exact algorithms for routing problems under vehicle capacity constraints.* Ann. Oper. Res., 75, 213–245.

Baldacci, R.; Bartolini, E.; Mingozzi, A. (2011). *An exact algorithm for the pickup and delivery problem with time windows*. Oper. Res., 59, 414–426

Beamon, B. (1998). *Supply Chain design and analysis: Models and methods*.

Cordeau, J.F.; Desaulniers, G.; Desrosiers, J.; Solomon, M.M.; Soumis, F. (2002). *The VRP with time windows. In The Vehicle Routing Problem;* Toth, P., Vigo, D., Eds.; SIAM Monograp: Philadenphia, PA, USA; pp. 157–186

Chen, X.; Kong, Y.; Dang, L.; Hou, Y.; Ye, X. (2015). *Exact and metaheuristic approaches for a biobjective school bus scheduling problem*. *PLoS ONE*, *10*, e0132600

Dantzig, G.B.; Ramser, J.H. (1959). *The truck dispatching problem. Manag. Sci.*, *6*, 80–91

G. Clarke and J. W. Wright, (1964) "*Scheduling of Vehicles from a Central Depot to a Number of Delivery Points*,".

J. Bramel and D. Simchi-levi, (1995), *"Operations Research,"*.

Jian, L.; Li, Y.; Panos, M.P. (2016). *Multiple-depot vehicle routing problem with time windows under shared depot resources.* J. Comb. Optim., 31, 515–532.

Laporte, G.; Nobert, Y. (1987). *Exact Algorithms for the Vehicle Routing Problem*. N.-Holl. Math. Stud. , 132, 147–184.

R. L. Bowerman and P. H. Calamai, (1994). *"Theory and Methodology the Spacefilling Curve with Optimal Partitioning Heuristic for the Vehicle Routing Problem,"* vol. 76, pp. 128–142.

MACHINE LEARNING IN MINIMIZING ERROR FOR FORECASTING NEW COVID-19 CASES IN UNITED STATES: LSTM, BILSTM, GRU

Dao Vu Truong Son[1] and Nguyen Minh Khoi[2]

[1] Department of Industrial and System Engineering at International University, Ho Chi Minh city, Vietnam, dvtson@hcmiu.edu.vn

[2] Department of Industrial and System Engineering at International University, Ho Chi Minh city, Vietnam, IELSIU17032@student.hcmiu.edu.vn

Abstract. This paper focuses on minimizing errors of forecasting new COVID-19 cases in United States by using machine learning methods. Forecasting is the process of predicting the future using the past and current data. However, the uncertainty in real-world data makes this process challenging. Moreover, demand for robust and precision in prediction is becoming higher than ever. Accurate prediction for future results can lead to a more efficient use of resources, manpower and better preparation for the upcoming events, thus increase revenue, maximize profit as well as reduce future risks of the business. The forecasting methods used are ARIMA, Exponential Smoothing, Gate Recurrent Unit (GRU), Long Short-term Memory (LSTM) and Bidirectional Long Short-term Memory (BiLSTM). The objective is to make a comparison and find out which methods provide the best result with the lowest set of error scores. The experimental results reveal that machine learning methods outperform the traditional methods, particularly BiLSTM have the best performance among five methods in terms of prediction accuracy, robustness and better understanding of context.

Keywords: Recurrent Neural Network; Time Series Forecasting; Coronavirus; BiLSTM; LSTM; GRU.

1. Introduction

Accurate time-series forecasting is critical for distinctive detection, optimal resources allocation, budget planning and specifically in this situation, better preparation and risk management. However, this problem is challenging because the event depends on numerous external factors, which can include movement, cities, interactions, weather, etc. The rising trend of COVID-19 can be learned in time through the prediction of the above-mentioned nonlinear time series, and then relevant personnel can take corresponding measures and strengthen the prevention and control for the future trend of the pandemic before it becomes serious.

Time series data often contain seasonality and trend, depends on different type of data. There are two ways of analyzing time series. One is referred to as fundamental analysis, and the other is called technical analysis. The research to develop a good forecasting model and to improve the effectiveness of existing forecasting models has been an active research area.
Traditional forecasting methods have proven their efficiency on linear assumption of a dataset, which can be pre-processed with noise, seasonality and trend in order to perform better. As the matter of fact, ARIMA and Simple Moving Average method are being used across corporations, which can act as a proof that their performances can help corporations achieve what they want. However, evidence have shown that when dealing with non-linear dataset, machine learning and deep learning methods outperforms the traditional methods due to lack of flexibility as well as the high level of complexity of the dataset. One would expect RNN, LSTM and its variation, Bidirectional LSTM, which are more advanced types of recurrent neural network, to be much more accurate than the ARIMA and other statistical methods. As we can see later, by calculating error scores, RNN, LSTM and Bidirectional LSTM provides more precise results compared to traditional methods.

The aim of this research is to forecast the number of covid-19 cases in United States and compared the results to the traditional methods to conclude which ones are worth taking into further consideration. By providing a large amount of data it was shown that BiLSTM approach can model complex nonlinear feature interactions, which is critical to model complex events. However, it is worth noting that the accuracy of the model may vary for different datasets. One can have better performance on datasets which includes trend or seasonality feature, or sometimes, both.

The purpose of this study is to minimize errors among different forecasting methods and find out which methods are better applied into dataset of Covid-19 cases in United States, which means that the dataset only lies in territorial boundary of USA and is not influenced by any external factors.

Some pros and cons can be seen in machine learning and deep learning methods in general, and they are:

- Machine learning can easily outperform traditional methods based on many error scores, which will be used in this work for comparison.
- Machine learning is capable of automatically learning and extracting feature from raw data, which means that it can be developed directly without the need of pre-processing, making the data stationary by differencing or scaling.

However, it still provides outcomes based on different scenarios such as multi-step forecasting, multivariate problems, different datasets, etc. Searching for hyper-parameters using grid search or random search is much more daunting and complex than that of traditional methods. Particularly, it will be more time-consuming, more values are taken into account, which are more prone to false and incorrect results.

2. Methodology for Coronavirus Prediction

This work explores the ability of each method on forecasting this specific type of data, therefore the initiate process will be the same for all of the research methods. Two kinds of methodologies will be examined, which is statistical models, including ARIMA and Triple Exponential Smoothing, and machine learning models, including LSTM, BiLSTM and GRU. The statistical metrics in terms of MAPE, SMAPE, MAE, RMSE are also specified in this section for performance evaluation. A dataset on confirmed Covid-19 cases in USA is collected and after preprocessing, data is passed to each model and their performances will be examined by error scores. As can be seen, more details of each model are provided below in this section.

2.1. Autoregressive Integrated Moving Average - ARIMA

ARIMA was first introduced in 1970 by Box and Jenkins. Up to now, ARIMA has shown efficient ability to generate short-term forecasts. The future values produced by ARIMA model is a linear combination of past values of data and past errors. This is due to the fact that its structure has the processed of auto regression, integration and moving average. ARIMA formula can be expressed as follows:

$$Z_t = \mu + \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \cdots + \phi_p(Z_{t-p} - \mu) + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}, \quad (1)$$

In Eq. (1), Z_t is the demand at time t, μ is the process average, and a_t is a time series of independent identically distributed random variables with expected. value $E(a_t) = 0$ and variance $V(a_t) = \sigma_a^2$.

2.2. Triple Exponential Smoothing - TES

Triple Exponential Smoothing is an advanced of Exponential Smoothing that explicitly adds support for seasonality feature to the univariate time series. This method is also named after two contributors: Holt-Winters Exponential Smoothing.

$$s_t = \alpha \frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \quad (2)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \quad (3)$$

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L} \quad (4)$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad (5)$$

Where in Eq. (2), y_t is the observation at time t, I_t is the seasonal index at time t, S_{t-1} is the smoothed statistic at previous time point, b_{t-1} is the estimated previous trend at time t and α is the smoothing factor for the level and it should be greater than 0 and less than 1, which value of 0 set the current smoothed point to the previous smoothed value and value of 1 sets the current smoothed point to the current point. In Eq. (3), b_t represents the estimated trend at time t, γ is the smoothing factor for the seasonality, S_t is the smoothed statistic at time t. β in Eq. (4) is the smoothing factor for the trend and its value is similar to α, The closer β is to 1, the less the prior data points enter into the smooth. Finally, F in Eq. (5) indicates the forecast value at m periods ahead.

As with the trend, the seasonality can have either value, an additive or multiplicative process. Triple exponential smoothing is the most advanced variation of exponential smoothing and through configuration. As an adaptive method, triple exponential smoothing allows the level, trend, and seasonality patterns to change throughout the dataset.

2.3. Long-Short Term Memory - LSTM

Vanishing gradient problem is one of the major problems of deep RNN. Therefore, LSTM or GRU are introduced as the more advance versions of RNN, have been used to overcome this drawback. Developed by Hochreiter and Schmidhuber, LSTM's structure consists of multiple gated cells that can allow data to pass through according to the import of the received data element. Every gated unit has a certain weight value estimated through the training phase using backpropagation operation. A LSTM unit combines a cell, an input gate, an output gate and a forget gate. The three gates are responsible for the flow of information into and out of the cell. The weight values in each gate represent thresholds by which data is stored or deleted.

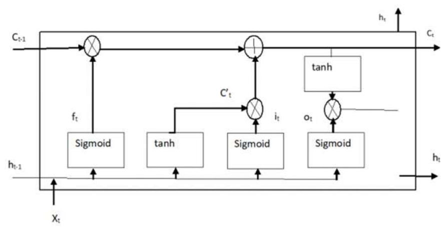


Figure 1. Structure of LSTM Cell

LSTM network is compute mapping between input sequence and output sequence: X = (X_1, X_2, …, X_n) and y = (y_1, y_2, …, y_n). Calculating by the following equations:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (6)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (7)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (8)$$

$$\tilde{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad (9)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \quad (10)$$

$$h_t = o_t \otimes \sigma_h(c_t) \quad (11)$$

In Eqs. (6), (7), (8), (9) and (10), W_f, W_o, W_C, W_i and b_f, b_i, b_o, b_c are the weights and bias variables respectively of three gates and a memory cell. Here, h_{t-1} symbolizes the previous hidden layer's units that element-wise adding with weights of three gates. After the processing of Eq. (10), C_t becomes current memory cell unit. Eq. (11) shows the element-wise multiplication of previous hidden unit outputs and previous memory cell unit. Tanh and sigmoid funtions are also added the non-linearity on top of three gates, which is shown in Eqs. (6–11). Here, t − 1 and t are prior and current time steps.

2.4. Bidirectional Long-Short Term Memory - BiLSTM

One limitation of LSTM can be pointed out is that it only uses the previous content, but do not have the ability to use the future one. Schuster and Paliwal proposed bidirectional recurrent neural networks (BRNN) that is comprised of two LSTM hidden layers which have opposite directions. With this architecture, previous and future content is fully exploited in the output layer. In BiLSTM, an input data sample X = (X_1, X_2, …, X_n) is calculated in forward manner $\overrightarrow{h_t} = (\overrightarrow{h_1}, \overrightarrow{h_2}, \dots, \overrightarrow{h_n})$ and backward manner $\overleftarrow{h_t} = (\overleftarrow{h_1}, \overleftarrow{h_2}, \dots, \overleftarrow{h_n})$. And the final output is formed by both $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$. The LSTM and biLSTM structure figures is displayed below.

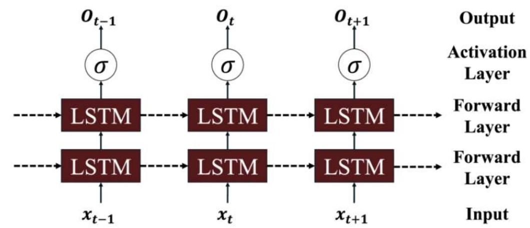


Figure 2. LSTM Network

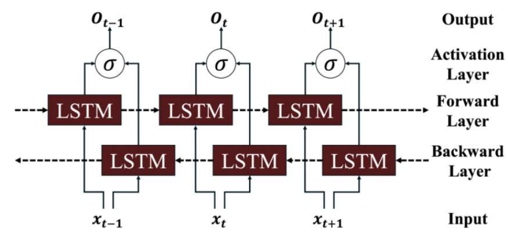


Figure 3. BiLSTM Network

2.5. Gated Recurrent Unit - GRU

GRU is an extended version of LSTM. In other words, it is the simple version of LSTM which has two gates, the update and reset gates. The update-gate is responsible of renewing the current memory of the network which enables the network to remember crucial data input of its import. The reset gate has responsibility of erasing the unusable memory of the network, which allows the network to forget certain values at designated time step.

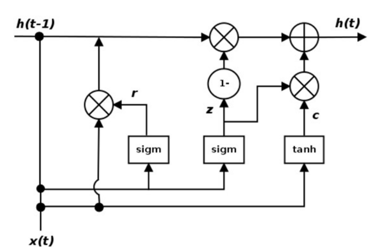


Figure 4. Structure of GRU Cell

GRU's performance is comparable to that of LSTM, but uses a smaller number of parameters, which makes it's more robust to train and less complex in structure. The equations of GRU hidden units are given as follows:

$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z) \quad (12)$$
$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r) \quad (13)$$
$$\hat{h}_t = \phi_h(W_h x_t + U_h(r_t \otimes h_{t-1}) + b_h) \quad (14)$$
$$h_t = (1 - z_t) \otimes h_{t-1} + z_t \otimes \hat{h}_t \quad (15)$$

Here, z_t, which is the update gate, in Eq. (15) decides for which kind of content or information is used and updated. In Eq. (14), the value of reset gate r_t is similar to update gate, if the gate is set to zero, it used the current input data points and forget the previously state. Further, $\hat{h}_t$ shows the same functionality as in recurrent unit and h_t of GRU at time t is the linear interpolation among the current $\hat{h}_t$ and previous h_{t-1} activation states Eq. (12) and (13).

2.6. Performance Metrics

The results of proposed model will be evaluated by four performances measures. These are Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Symmetric Mean Absolute Percentage Error (SMAPE). Four metrics are expressed mathematically as below, beginning in Eq. (16) is MAPE:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (16)$$

where A_t is the actual value and F_t is the forecast value. SMAPE is also defined in Eq. (17), with similar notation as MAPE as:

$$SMAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \quad (17)$$

Calculation for RMSE is displayed in Eq. (18):

$$RMSE(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E((\hat{\theta} - \theta)^2)} \quad (18)$$

RMSE is always non-negative, and a value of 0 would indicate a perfect fit to the data. Lastly, formula for MAE is shown in Eq. (19):

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n} \quad (19)$$

3. Experimental Results and Analysis

To investigate the proposed methodologies on coronavirus data, we use historical data provided by WHO. The .csv file of confirmed cases and the dates of United State is provided. COVID-19 dataset contains number of confirmed cases and we have used cases from the beginning of 2020, 1/22/2020 to 10/12/2020 for training and from 11/12/2020 to 2/18/2021 for testing. The data is pre-processed before it is given to ML models for training. All developed models are implemented in Python using Keras and TensorFlow libraries. Unscaled data slows down the convergence process, therefore we use MinMaxScaler to help with data shape problem. It does not change the meaning of the information in the original data and also does not decrease the instrumental features of outliers. Below are the parameter values used for the machine learning models. It is worth noticed here that parameters of all using methods has been carefully carried out through trial and error and values enlisted in Table 1.

Table 1. Value of Parameters for Neural Networks

	LSTM	BiLSTM	GRU
Hidden Layers	1	1	1
Hidden Neurons	128	128	128
Gate Activation	Sigmoid	Sigmoid	Sigmoid
Activation	Tanh	Tanh	Tanh
# Of Epochs	100	100	100
Optimizer	ADAM	ADAM	ADAM
Dropout	20%	20%	20%
Batch Size	32	32	32

3.1. Result of ARIMA for COVID-19 dataset

Table 2 is the result of the predicted values of ARIMA (0, 1, 1) considered the best model for confirmed coronavirus cases in USA. Figure 5 gives graphical results of the precision of the predicted cases against actual confirmed cases to see the performance of the ARIMA model.

Table 2. Numerical Result of ARIMA

Sample Period	Actual Values	Predicted Value
1/10/2021	213281	266345.9122
1/11/2021	213963	244050.5833
1/12/2021	226238	231485.4998
1/13/2021	229871	229439.4853
1/14/2021	234975	229798.3719
1/15/2021	240707	232166.6791
1/16/2021	200764	235959.4146
1/17/2021	177404	221231.3098
1/18/2021	142605	202847.8428
1/19/2021	176984	177512.8644
1/20/2021	182579	177465.0578
1/21/2021	193055	179806.8193
1/22/2021	190032	185593.1995
1/23/2021	170138	187649.0577
1/24/2021	131075	180409.7751
1/25/2021	151095	159694.0532
1/26/2021	146597	156228.7511
1/27/2021	152478	152326.1308
1/28/2021	168620	152566.5955
1/29/2021	166113	159540.9082
1/30/2021	142091	162500.1546
1/31/2021	111896	154033.6104
2/1/2021	134339	136365.6831
2/2/2021	114437	135683.5928
2/3/2021	121469	126862.4174
2/4/2021	123188	124754.6109
2/5/2021	133558	124267.3482
2/6/2021	104015	128377.8274
2/7/2021	89581	118237.0152
2/8/2021	89727	106278.1601
2/9/2021	95360	99445.36686
2/10/2021	94704	107891.4826
2/11/2021	105353	96717.82677
2/12/2021	99511	100550.7295
2/13/2021	83321	100286.586
2/14/2021	64938	93278.29521
2/15/2021	53944	81453.13851
2/16/2021	62398	69979.95303
2/17/2021	70188	66945.36395
2/18/2021	69228	68494.68052

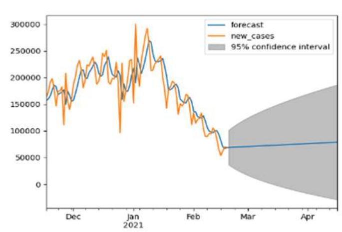


Figure 5. Graphical Result of ARIMA

3.2 Result of Triple Exponential Smoothing for COVID-19 dataset

Table 3 contained the predicted values of the Triple Exponential Smoothing model selected and figure 6 is the graph of predicted confirmed cases against confirmed cases to demonstrate the accuracy.

Table 3. Numerical Result of Exponential Smoothing

Sample Period	Actual Values	Predicted Value
1/10/2021	213281	223017.1981
1/11/2021	213963	228905.6527
1/12/2021	226238	238504.3514
1/13/2021	229871	225671.343
1/14/2021	234975	214896.8704
1/15/2021	240707	210877.8417
1/16/2021	200764	207148.9238
1/17/2021	177404	214009.3845
1/18/2021	142605	205980.7027
1/19/2021	176984	173247.1479
1/20/2021	182579	191539.0695
1/21/2021	193055	179731.4988
1/22/2021	190032	179083.4151
1/23/2021	170138	185358.7424
1/24/2021	131075	199410.4464
1/25/2021	151095	209954.7336
1/26/2021	146597	194545.6032
1/27/2021	152478	229696.6925
1/28/2021	168620	224262.2051
1/29/2021	166113	206543.3624
1/30/2021	142091	223064.1432
1/31/2021	111896	231396.5118
2/1/2021	134339	245945.037
2/2/2021	114437	264837.3302
2/3/2021	121469	267033.4119
2/4/2021	123188	269556.5643
2/5/2021	133558	271148.8569
2/6/2021	104015	271546.6517
2/7/2021	89581	271741.5744
2/8/2021	89727	271656.9778
2/9/2021	95360	274357.9277
2/10/2021	94704	277915.2149
2/11/2021	105353	278854.0984
2/12/2021	99511	280617.9575
2/13/2021	83321	281368.8181
2/14/2021	64938	280994.0623
2/15/2021	53944	282399.7118
2/16/2021	62398	283485.2839
2/17/2021	70188	287147.8585
2/18/2021	69228	290420.6048

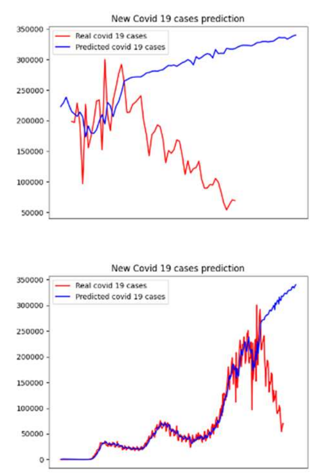


Figure 6. Graphical Result of Exponential Smoothing

3.3 Result of Gated Recurrent Unit for COVID-19 dataset

Similar characteristics of the model are also applied to GRU to produce the below results. Below displayed the graphical and numerical results for GRU. Although it suffers from minor overfitting problem, GRU can also be able to follow the downtrend of the data sample.

Table 4. Numerical Result of Gated Recurrent Unit

Sample Period	Actual Values	Predicted Value
1/10/2021	213281	221770.78
1/11/2021	213963	193263.34
1/12/2021	226238	193273.95
1/13/2021	229871	198637.5
1/14/2021	234975	199447.28
1/15/2021	240707	202144.38
1/16/2021	200764	205765.86
1/17/2021	177404	180678.84
1/18/2021	142605	165452.88
1/19/2021	176984	141214.22
1/20/2021	182579	157601.48
1/21/2021	193055	160372.66
1/22/2021	190032	166700.7
1/23/2021	170138	165555.78
1/24/2021	131075	153787.38
1/25/2021	151095	129036.6
1/26/2021	146597	137318.28
1/27/2021	152478	133927.4
1/28/2021	168620	136342.17
1/29/2021	166113	146069.92
1/30/2021	142091	145818.58
1/31/2021	111896	131802.36
2/1/2021	134339	112079.73
2/2/2021	114437	121574.336
2/3/2021	121469	109818.15
2/4/2021	123188	111463.61
2/5/2021	133558	112032.805
2/6/2021	104015	118031.78
2/7/2021	89581	101448.76
2/8/2021	89727	90126.42
2/9/2021	95360	87141.18
2/10/2021	94704	88783.766
2/11/2021	105353	88065.24
2/12/2021	99511	93863.83
2/13/2021	83321	91633.02
2/14/2021	64938	81968.04
2/15/2021	53944	69208.28
2/16/2021	62398	59322.477
2/17/2021	70188	60654.887
2/18/2021	69228	64599.336

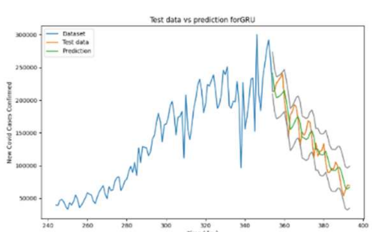


Figure 7. Graphical Result of Gated Recurrent Unit

3.4 Result of Long-Short Term Memory for COVID-19 dataset

This section also contains a graph with actual values and predicted values side by side, along with a table contained numerical values of the two. From the graph, the performance of the LSTM model selected is quite impressive.

Table 5. Numerical Result of Long-Short Term Memory

Sample Period	Actual Values	Predicted Value
1/10/2021	213281	274197.7408
1/11/2021	213963	257696.0382
1/12/2021	226238	232647.7207
1/13/2021	229871	231488.3668
1/14/2021	234975	235479.6183
1/15/2021	240707	238909.843
1/16/2021	200764	242925.2959
1/17/2021	177404	229804.5486
1/18/2021	142605	212229.3534
1/19/2021	176984	190691.5676
1/20/2021	182579	169562.0803
1/21/2021	193055	185476.0741
1/22/2021	190032	193481.0259
1/23/2021	170138	195534.1548
1/24/2021	131075	188299.965
1/25/2021	151095	172186.534
1/26/2021	146597	153874.0528
1/27/2021	152478	156884.1832
1/28/2021	168620	158442.4629
1/29/2021	166113	166729.3511
1/30/2021	142091	169384.7665
1/31/2021	111896	160799.3256
2/1/2021	134339	145595.6739
2/2/2021	114437	134210.1899
2/3/2021	121469	130975.4145
2/4/2021	123188	128877.2185
2/5/2021	133558	129652.1408
2/6/2021	104015	134360.1894
2/7/2021	89581	124661.7986
2/8/2021	89727	112522.3019
2/9/2021	95360	102815.3067
2/10/2021	94704	101863.6336
2/11/2021	105353	101918.0829
2/12/2021	99511	106139.9903
2/13/2021	83321	105703.0812
2/14/2021	64938	98642.40513
2/15/2021	53944	87148.57185
2/16/2021	62398	74414.63419
2/17/2021	70188	69120.41236
2/18/2021	69228	72906.60559

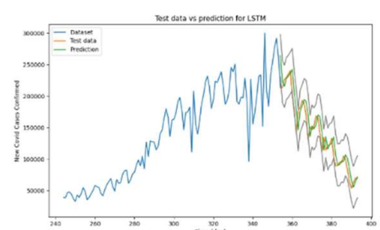


Figure 8. Graphical Result of Long-Short Term Memory

3.5 Result of Bidirectional Long-Short Term Memory for COVID-19 dataset

This section provides results of BiLSTM model, both graphical and numerical results. As can be seen, BiSTLM can be able to predict the patterns and the trend of the data sample well. Unlike other models, not only BiLSTM seem not to suffer from overfitting problem, but it also has the smallest error scores, which will be displayed below. From the graph, it is obvious that the performance is satisfactory.

Table 6. Numerical Result of Bidirectional Long-Short Term Memory

Sample Period	Actual Values	Predicted Value
1/10/2021	213281	196559.67
1/11/2021	213963	184333.92
1/12/2021	226238	181733.56
1/13/2021	229871	186204.78
1/14/2021	234975	191079.53
1/15/2021	240707	194372.28
1/16/2021	200764	195693
1/17/2021	177404	187714.23
1/18/2021	142605	179593.92
1/19/2021	176984	168813.16
1/20/2021	182579	167518.84
1/21/2021	193055	162877.02
1/22/2021	190032	158065.5
1/23/2021	170138	152594.8
1/24/2021	131075	149168.61
1/25/2021	151095	140905.53
1/26/2021	146597	139004.61
1/27/2021	152478	138506.88
1/28/2021	168620	137449.75
1/29/2021	166113	136423.94
1/30/2021	142091	131864.05
1/31/2021	111896	124205.24
2/1/2021	134339	112138.64
2/2/2021	114437	107971.69
2/3/2021	121469	106607.62
2/4/2021	123188	110071.164
2/5/2021	133558	110844.67
2/6/2021	104015	111332.4
2/7/2021	89581	101702.4
2/8/2021	89727	92076.78
2/9/2021	95360	89143.11
2/10/2021	94704	91812.27
2/11/2021	105353	93530.53
2/12/2021	99511	95729.71
2/13/2021	83321	93043.79
2/14/2021	64938	84629.086
2/15/2021	53944	73957.234
2/16/2021	62398	66333.03
2/17/2021	70188	65860.91
2/18/2021	69228	68014.45

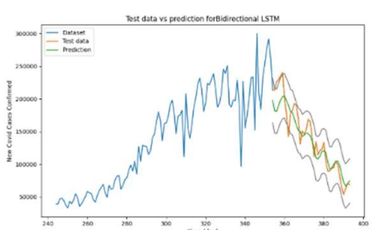


Figure 9. Graphical Result of Bidirectional Long-Short Term Memory

3.6 Performance Metrics Comparison

In this section, a summary table contains a set of performance metrics is displayed. This helps provide additional insights to which methods are performed best. A deeper examination among 4 metrics: RMSE, MAPE, SMAPE and MAE.

Table 7. Performance Metrics Summary

	MAE	RMSE	MAPE	SMAPE
ARIMA	61682.59	76026.65	53.86	1705.73
Exponential Smoothing	54578.81	63966.77	52.14	38.00
LSTM	15036.28	18915.95	11.50	11.10
BiLSTM	**14394.83**	**18203.01**	**11.26**	**10.88**
GRU	15495.68	18821.53	11.77	11.42

As can be seen, machine learning methods are heavily outperformed traditional methods. It can be observed from this table and predicted values tables above that none of the two models, ARIMA and Exponential Smoothing does not generate consistent predictions, they completely lack of flexibility and ability to follow when number of cases tend to increase or decrease suddenly. On the other hand, all the machine learning methods tend to capture the trends and extreme values, which results in a large difference in value of performance metrics. As the results, BiLSTM models produce better performance compared to the regular LSTMs. It seems that BiLSTMs are able to understand the context better by transferring inputs data twice (from left to right and vice versa). These plots provide an outstanding demonstration of the match of predicted cases against actual ones for all three techniques with much better performance than baseline benchmarks. Furthermore, among deep learning models, Bi-LSTM performs very well since it has the ability to learn from both ends of the data. Therefore, it seems to adapt and understand the context better than others. Our results show that BiLSTMs perform better compared to regular LSTMs and GRU in the context of predicting time series data.

3.7 Loss Function

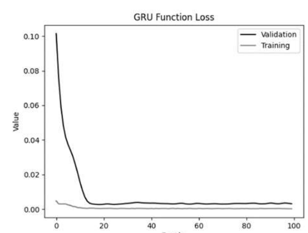


Figure 10. Loss Function of Gated Recurrent Unit

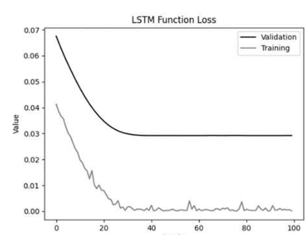


Figure 11. Loss Function of Long-Short Term Memory

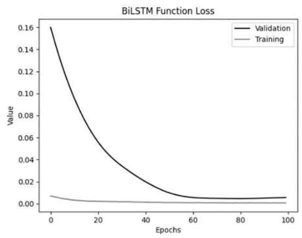


Figure 12. Loss Function of Bidirectional Long-Short Term Memory

According to the graphs, loss function has a clear downtrend as long as the training continues. This indicates that three machine learning models are learning correctly followed the data sample. Another conclusion can be drawn from the graphs is that the training loss is smaller than the validation loss, which indicates that the training sample is well-learned and adapted.

4. Conclusion and Further Work

This paper examined the results of an experiment, through which the performance and accuracy as well as behavioural training of ARIMA, Exponential Smoothing, Long Short-term Memory (LSTM), Gated Recurrent Unit (GRU) and bidirectional Long Short-term Memory (BiLSTM) models were analysed and compared. LSTM, GRU and Bi-LSTM have shown robustness and much advanced predictions when compared the performance errors, nevertheless, Bi-LSTM outperformed among all four models on the basis of four error measures. A conclusion can be drawn that Bi-LSTM is a suitable method for such time series data and capable of predicting with improved accuracy for other datasets which contain similar features for better management. We also noticed that training based on BiLSTM is slightly slower and it takes a certain number of additional batches of dataset to reach the equilibrium state. This observation means that there are some features associated with dataset that might be exposed by BiLSTM but unidirectional LSTM cannot capture, due to the fact that the training is only one way (i.e., from left to right). All things considered, this paper highly recommends using BiLSTM instead of LSTM and GRU for forecasting problems in time series analysis. One should consider the result of this research to tackle the problem of multivariate time series problem. in the real world, an event can be affected by many factors.

Reference

Ahmed, Nesreen K., et al. (2010). "An empirical comparison of machine learning models for time series forecasting." Econometric Reviews 29.5-6: 594-621.

Ariyo, Adebiyi A., Adewumi O. Adewumi, and Charles K. Ayo. (2014). "Stock price prediction using the ARIMA model." 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE.

Billah, Baki, et al. (2006). "Exponential smoothing model selection for forecasting." International journal of forecasting 22.2, 239-247.

Carbonneau, Real, Kevin Laframboise, and Rustam Vahidov. (2008). "Application of machine learning techniques for supply chain demand forecasting." European Journal of Operational Research 184.3, 1140-1154.

Ding, Daizong, et al. (2019). "Modeling extreme events in time series prediction." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Dey, Rahul, and Fathi M. Salem. (2017). "Gate-variants of gated recurrent unit (GRU) neural networks." IEEE 60th international midwest symposium on circuits and systems (MWSCAS). IEEE, 2017.

Eade, Rosie, et al. (2012). "Forecasting the number of extreme daily events out to a decade ahead." Journal of Geophysical Research: Atmospheres 117.D21.

Fu, Rui, Zuo Zhang, and Li Li. (2016). "Using LSTM and GRU neural network methods for traffic flow prediction." 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 2016.

Gardner Jr, Everette S. (1985). "Exponential smoothing: The state of the art." Journal of forecasting 4.1 (1985): 1-28.

Global Change Data Lab Project, a registered charity. https://ourworldindata.org/coronavirus-source-data

Goodwin, Paul, and George Wright. (2010). "The limits of forecasting methods in anticipating rare events." Technological forecasting and social change 77.3: 355-368.

Hillmer, Steven Craig, and George C. Tiao. (1982). "An ARIMA-model-based approach to seasonal adjustment." Journal of the American Statistical Association 77.377: 63-70.

Ho, Sui-Lau, Min Xie, and Thong Ngee Goh. (2002). "A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction." Computers & Industrial Engineering 42.2-4: 371-375.

Hoerling, Martin, et al. (2013). "Anatomy of an extreme event." Journal of Climate 26.9: 2811-2832.

Hochreiter, Sepp, and Jürgen Schmidhuber. (1997). "Long short-term memory." Neural computation 9.8: 1735-1780.

Kong, Weicong, et al. (2017). "Short-term residential load forecasting based on LSTM recurrent neural network." IEEE Transactions on Smart Grid 10.1 : 841-851.

Kumar, D. Nagesh, K. Srinivasa Raju, and T. Sathish. (2004). "River flow forecasting using recurrent neural networks." Water resources management 18.2: 143-161.

Minh, Dang Lien, et al. (2018). "Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network." Ieee Access 6 (2018): 55392-55404.

Muzaffar, Shahzad, and Afshin Afshari. (2019). "Short-term load forecast using LSTM networks." Energy Procedia 158 (2019): 2922-2927.

Nabi, Khondoker Nazmoon. (2020). "Forecasting COVID-19 pandemic: A data-driven analysis." Chaos, Solitons & Fractals 139 (2020): 110046.

Qi, Di, and Andrew J. Majda. "Using machine learning to predict extreme events in complex systems." Proceedings of the National Academy of Sciences 117.1: 52-59.

Sarkar, Kankan, Subhas Khajanchi, and Juan J. Nieto. (2020). "Modeling and forecasting the COVID-19 pandemic in India." Chaos, Solitons & Fractals 139: 110049.

Siami-Namini, Sima, Neda Tavakoli, and Akbar Siami Namin. (2019). "The performance of LSTM and BiLSTM in forecasting time series." 2019 IEEE International Conference on Big Data (Big Data). IEEE.

Shen, Guizhu, et al. (2018). "Deep learning with gated recurrent unit networks for financial sequence predictions." Procedia computer science 131: 895-903.

Shepherd, Theodore G. (2016). "A common framework for approaches to extreme event attribution." Current Climate Change Reports 2.1: 28-38.

Schuster, Mike, and Kuldip K. Paliwal. (1997) "Bidirectional recurrent neural networks." IEEE transactions on Signal Processing 45.11: 2673-2681.

Vermaak, J., and E. C. Botha. (1998). "Recurrent neural networks for short-term load forecasting." IEEE Transactions on Power Systems 13.1: 126-132.

Yu, Rose, et al. (2017). "Deep learning: A generic approach for extreme condition traffic forecasting." Proceedings of the 2017 SIAM international Conference on Data Mining. Society for Industrial and Applied Mathematics.

Zhao, Zheng, et al. (2017). "LSTM network: a deep learning approach for short-term traffic forecast." IET Intelligent Transport Systems 11.2: 68-75.

Zhang, G. Peter, and Min Qi. (2005). "Neural network forecasting for seasonal and trend time series." European journal of operational research 160.2: 501-514.

Zhang, Xun, et al. (2009). "Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method." Energy Economics 31.5: 768-778.

Zhu, Lingxue, and Nikolay Laptev. (2017). "Deep and confident prediction for time series at uber." IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2017.

AN APPLICATION OF DATA ENVELOPMENT ANALYSIS TO EVALUATE THE PROFITABILITY AND MARKETABILITY EFFICIENCY OF PHARMACEUTICAL COMPANIES IN VIETNAM

Xuan-Huynh Nguyen[1], Quoc-Tuan Nguyen[2], and Thi-Kim-Lien Nguyen[3]*

[1] Hanoi School of Business & Management, Vietnam National University, Hanoi 100000, Vietnam

[2] Faculty of Tourism, Hoa Binh University, Hanoi 100000, Vietnam

[3] Scientific Research-International Cooperation, Thanh Dong University, Hai Duong 171967, Vietnam

Abstract. The pharmaceutical industry is an essential aspect of human life around the world because it researches and provides medicine. In this study, we investigate the performance of the pharmaceutical industry in Vietnam through the negative Malmquist model in Data Envelopment Analysis (DEA). This study applies the Malmquist model with the presence of negative values in DEA to evaluate the effectiveness of listed Vietnamese pharmaceutical companies from Q1/2019 to Q2/2020. The empirical results showed efficiency in a two-stage model, inclusive of business performance and market performance, through the negative Malmquist Index as it changes from time-to-time. The analysis results revealed that in the first stage, inputs such as operating costs and equity were those factors most directly influencing business performance; whereas, in the second stage, output values in the first stage were the input values for stage two, which were the factors most directly affecting market efficiency. The research results, thus, gave a performance overview of each pharmaceutical company in each period. Moreover, the research also informed the solution to improve overall efficiency in the phases where good efficiency did not occur

Keywords: Data Envelopment Analysis – DEA; Negative Malmquist model; efficiency; pharmaceutical industry

1. Introduction

The development of the pharmaceutical industry has always been an important factor in the human therapeutic process because humans have used drugs to treat illness and disease for more than 3000 years (David, 2016). The function of a pharmaceutical company is to research, develop, market, and distribute drugs. Although the global pharmaceutical industry was found in China about 1100 BCE (Leung, 2006), at the end of 19 century, it was actually established and developed sharply. From the earlier of 2020 to now, the pharmaceutical companies over the world are facing the impact of COVID-19, the pandemic has reduced the number of novel late-stage assets (Elmhirst & Urquhart, 2020). In recent years, the pharmaceutical industry in Vietnam has experienced rapid development. The COVID-19 pandemic has negatively impacted both the domestic and international pharmaceutical industry continuously since early 2020. To gain an overview of development in the pharmaceutical industry in Vietnam, both before and after the pandemic, this study utilized the negative Malmquist model in DEA to provide an overall evaluation of business activities and marketability for twelve pharmaceutical companies listed on the Vietnamese Stock Market since the beginning of 2019 to the end of Q2/2020. The purpose of this study is to evaluate the efficiency of pharmaceutical companies in Vietnam from 2019 to the end of Q2/2020, the collected data was analysed in two stages. Stage one analyzed the early input variables, including total assets, operating costs, and equity to provide output variables which are net revenue from sales and service providers to indicate business efficiency. The second stage was to analyzed market performance using output values available from stage one as input variables to capture output variables such as earnings per share and book-value of shares. The empirical results found out the business efficiency and market efficiency of each pharmaceutical company in each period.

Data envelopment analysis (DEA) has been a useful tool since it includes many models, such as the efficiency Slacks-Based Model (SBM), higher efficiency model (Super-SBM), etc. DEA is a non-parametric method used in economics and operational research to calculate levels of efficiency. It helps to provide a measure for the efficiency of a specific Decision-Making Unit (DMU). Results of the analysis provide benchmarks to select efficiency methods in the production and business process. The first traditional model was brought about by Charnes, Cooper, and Rhodes (1978) (Charnes, Cooper, & Rhodes, 1978) and became known as the CCR model. Since that time, many researchers built and developed newer models with diverse resolution for variables. This included the undesirable outputs model (Seiford & Zhu, 2002) developed to simultaneously deal with desired output variables and unwanted output variables. In the early stages of development, the DEA model could only give the maximum efficiency score of 1. Tone (2002), however, has produced the basic theory of unlimited efficiency scores in which scores are calculated based on available input and output variables. Another common feature of DEA is its use of a non-parametric method and a linear program.

The Malmquist model introduced by Cave, Christensen, and Diewert (1982) shows the change in efficiency over time. Fare, Grosskopf, Norris, and Zhang (1992) further developed it and included it in DEA. The Malmquist model offers technical efficiency for each period, combined with high efficiency, to give an unlimited maximum score. The Malmquist model delivers values based on technical changes and technical efficiency. According to Aghayi, Tavana, and Maleki, (2019), technical change and technical efficiency change is the sum of the factors that change efficiency and are measured by the Malmquist Index (MI). MI helps to identify past unit operations and to offer a new direction in the future. In addition, the Malmquist model was developed to solve negative values (Ghasem, Shabnam, & Simin, 2014). With these characteristics, many researchers have chosen the negative Malmquist model to measure the performance in various fields such as banking (Maria & Emmanuel, 2010), economics (Esmaeili & Malkhalifeh, 2017), industry (Naser, Hadi, & Ali, 2018), and environmental engineering (An, Wu, Li, Xiong, & Chen, 2019). The negative Malmquist model, as well as the data inclusion method, have been utilized by researchers in different fields to measure the structured performance in that industry using selected input and output variables. It should be noted, however, depending on the profession involved, the research goals, the number of variables, and types of variables are selected accordingly to meet particular needs.

In the early stages of formation, the Malmquist model can only work utilizing positive values, but in the following research efforts, it has developed and resolved the presence of negative values in the original data. Therefore, this study used the negative Malmquist model in DEA to deal with the presence of negative data and measure the performance of pharmaceutical companies in Vietnam. This paper introduces the negative Malmquist model in DEA to work with negative values based on previous research carried out by Esmaeili and Malkhalifeh (2017). Previous models set out to evaluate efficiency were only able to calculate positive values and eliminate the negative values. The strength of this method, thus, is to provide a comprehensive view and insight into the pharmaceutical industry in Vietnam in recent periods regardless of valuation. The research results reflect the current situation of Vietnam's pharmaceutical industry. Each company can re-examine its relative advantages and disadvantages and draw future development directions. In addition, the research reveals a detailed evaluation in each phase based on the Malmquist model through two stages: efficiency in business and marketability efficiency.

The rest of the study is arranged in four sections as follows. Section 1 overviewed the pharmaceutical industry, data envelopment analysis, and the negative Malmquist model. Section 2 showed the data source, proposal research, and formula equations of the negative Malmquist model. Section 3 conducted the analysis results of business efficiency and market efficiency. Section 4 summarized the main results which were figured out.

2. Materials and method

2.1. Data collection

Based on the given research objectives and models, data obtained from 12 pharmaceutical companies listed on the Vietnam stock market since the beginning of 2019 to the end of Q2 2020 was collected from website: https://vietstock.vn/ (Vietstock, 2020). The list of pharmaceutical companies is shown in Table 1.

Table 1. List of 12 pharmaceutical companies in Vietnam

Symbol	Company names
DBD	Binh Dinh Pharmaceutical - Medical Equipment Joint Stock Company
DCL	Cuu Long Pharmaceutical Joint Stock Company
DHG	Hau Giang Pharmaceutical Joint Stock Company
DMC	Domesco Medical Import-Export Joint Stock Company
DP3	Central Pharmaceutical Joint Stock Company 3
IMP	Imexpharm Pharmaceutical Joint Stock Company
OPC	OPC Pharmaceutical Joint Stock Company
PME	Pymepharco JSC
PPP	Phong Phu Pharmaceutical Joint Stock Company
SPM	SPM JSC
TRA	Traphaco JSC
VDP	Central Pharmaceutical Joint Stock Company VIDIPHA

With the research objectives and principles of MI in the DEA method, the input and output variables from the financial reports were selected particularly.

Total assets (TAS): The total amount of assets includes tangible and intangible which are owned by the pharmaceutical company.

Operating expenses (OES): The operating expenses are rent, equipment, inventory cost, marketing, payroll, insurance, step costs, and funds.

Owner’s equity (QEY): The owner’s equity is the owner’s investment in the business.

Net revenue (NRE): The net revenue is the sale of a pharmaceutical company when this value is returned from the selling operation.

Profit after tax (PAT): The profit after tax is the earnings of a business when the total revenue deducts all income taxes.

Earnings-per-share (EPS): The earnings-per-share of a pharmaceutical company is a part of net profit, excluding extraordinary items or discontinued operations, or on a diluted basis.

Book value per share (BVPS): The book value per share is the ratio of equity available to common shareholders.

2.2. Proposal research

The proposal research of pharmaceutical company is drawn in the Figure 1.

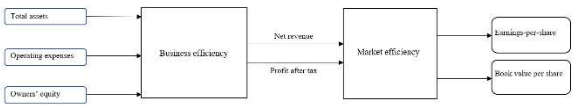


Figure 1. Proposal research

Figure 1 showed that the research is divided into two stages. In the first stage which evaluated business performance, the selected input variables included: Total assets, operating expenses, and owners' equity; the output variables included: Net revenue and profit after tax, these factors were used to measure the performance of the business process. In the second stage of market efficiency, the selected input variables were given as the output variables in stage one, including net revenue and profit after tax; the output variables included earnings-per-share and book value per share, the market efficiency of each pharmaceutical company in each period was conducted.

With the appearance of negative data in the outputs of stage 1 and inputs of stage 2, the negative Malmquist model in DEA is used to measure the efficiency of business process and market process. Each score of these pharmaceutical companies in each stage and each period is calculated based on the historical data.

2.3. Malmquist model

The Malmquist model is used to measure the performance of DMU's during successive periods, it evaluates the efficiency change over time with the positive data (Tone, 2016). In the previous time, Saitech group provided DEA-Solver-PRO versions without dealing with the negative data, the current version – DEA-Solver-PRO provided the fifteen version in which solves with the negative data. Tohidi, Razavyan, and Tohidnia, (2014) introduced the presence of the negative data in DEA. Thus, this paper used the Malmquist model with negative data to provide an analysis of the efficiency of pharmaceutical companies located in Vietnam in recent times. There scores are calculated according to the negative Malmquist model in DEA. The negative Malmquist model, thus, evaluates the value by the ratio of total output to total input in each period. Each DMU is set with input and output variables as x^t, y^t respectively. The distance between set points is given as D. The MI score at each point is listed as $D_0^t(x_0^t, y_0^t)$, $D_0^{t+1}(x_0^t, y_0^t)$, $D_0^t(x_0^{t+1}, y_0^{t+1})$, and $D_0^{t+1}(x_0^{t+1}, y_0^{t+1})$.

Catch-up efficiency is calculated by:

$$Catch-up = \frac{D_o^{t+1}(x_o^t, y_o^t)}{D_o^t(x_o^t, y_o^t)} \quad (1)$$

Frontier-shift (FS) efficiency is conducted by:

$$FS = \left[\frac{D_o^t(x_o^{t+1}, y_o^{t+1})}{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})} \times \frac{D_o^t(x_o^t, y_o^t)}{D_o^{t+1}(x_o^t, y_o^t)} \right]^{1/2} \quad (2)$$

To combine equation (1) and (2), the equation of Malmquist model is formulated as follows:

$$Malmquist = \left[\frac{D_0^t(x_0^{t+1}, y_0^{t+1})}{D_0^t(x_0^t, y_0^t)} \times \frac{D_0^{t+1}(x_0^{t+1}, y_0^{t+1})}{D_0^{t+1}(x_0^t, y_0^t)} \right]^{1/2} \quad (3)$$

Table 2. Data description of 12 farmaceutical companies

Indicator	Time	TAS	OES	QEY	NRE	PAT	EPS	BVPS
Max	Q1/2019	4,234,594	212,046	3,017,944	767,191	135,174	11,454	54,499
Min		146,760	5,297	96,948	30,503	-2,688	62	12,119
Average		1,456,267	69,665	1,054,342	264,883	36,092	3,854	28,110
SD		1,001,251	56,255	762,470	185,872	36,385	2,894	11,563
Max	Q2/2019	4,335,180	286,430	3,054,381	975,823	172,919	11212	54,570
Min		148,516	5,825	98,146	40,740	-2,479	-9	12,268
Average		1,476,260	84,254	1,044,543	305,557	40,144	3,827	27,120
SD		1,034,583	83,148	761,771	239,123	46,389	2,850	10,569
Max	Q3/2019	3,927,167	266,665	3,173,706	874,357	117,477	9,902	54,679
Min		150,497	6,035	102,489	39,283	1,776	138	11,647

Indicator	Time	TAS	OES	QEY	NRE	PAT	EPS	BVPS
Average		1,474,775	82,987	1,079,628	299,006	36,102	3,752	27,954
SD		947,490	69,860	792,013	211,626	32,386	2,545	10,800
Max	Q4/2019	4,146,819	358,361	3,377,551	1,279,383	203,845	10,421	55,001
Min		157,942	7,372	106,058	41,003	4,023	680	12,052
Average		1,516,127	97,259	1,113,238	382,436	54,245	3,932	28,476
SD		995,719	97,060	846,072	318,246	54,851	2,503	10,889
Max	Q1/2020	4,194,259	260,661	3,529,584	858,456	176,970	10,084	55,159
Min		153,036	6,617	107,784	29,584	1,949	690	12,248
Average		1,511,905	85,137	1,150,654	298,124	41,458	3,974	29,265
SD		1,005,530	69,052	883,260	210,975	45,680	2,401	10,978
Max	Q2/2020	4,297,132	254,312	3,192,212	820,292	185,736	9,921	54,307
Min		151,521	5,965	103,330	27,465	749	673	11,742
Average		1,537,704	69,351	1,104,389	275,395	45,197	4,047	27,711
SD		1,027,219	69,060	798,849	204,948	47,284	2,352	10,516

Table 3. Business efficiency

DMUs	Q1/2019	Q2/2019	Q3/2019	Q4/2019	Q1/2020	Q2/2020	Average
DBD	0.9140	0.9704	0.9528	1.3165	1.1808	0.9458	1.0467
DCL	0.5557	0.5859	0.5799	1.9289	0.6372	0.6382	0.8210
DHG	1.0000	1.0821	1.0386	1.1792	1.1939	1.1939	1.1146
DMC	1.2808	1.7241	1.5631	1.5244	1.2072	1.7609	1.5101
DP3	1.5711	3.7661	1.4931	0.8602	1.3671	2.3611	1.9031
IMP	0.6520	0.6947	0.7416	1.1078	0.7213	0.8557	0.7955
OPC	1.0413	1.1311	1.1532	0.9477	1.1129	0.8504	1.0394
PME	0.8402	1.0092	0.8886	2.0625	1.0308	0.9086	1.1233
PPP	3.8429	3.1953	4.2150	3.9778	3.6439	3.4550	3.7216
SPM	0.5915	0.4904	0.6503	0.6717	0.4562	0.5220	0.5637
TRA	1.1250	1.3591	0.9577	1.5082	1.0935	1.3586	1.2337
VDP	1.2207	1.2164	1.1711	1.1974	1.1364	0.9470	1.1482

Table 4. Market efficiency

DMUs	Q1/2019	Q2/2019	Q3/2019	Q4/2019	Q1/2020	Q2/2020	Average
DBD	0.2058	0.1790	0.1876	0.1549	0.1448	0.1994	0.1786
DCL	8.4701	4.0826	0.4006	0.1110	0.6143	0.4582	2.3561
DHG	0.0803	0.0654	0.0848	0.0550	0.0779	0.0829	0.0744
DMC	0.2749	0.2347	0.2352	0.2152	0.3587	0.2950	0.2689
DP3	1.0000	1.0261	1.1152	1.4598	1.4598	2.0378	1.3498
IMP	0.2355	0.2489	0.2169	0.1567	0.2168	0.2017	0.2127
OPC	0.2923	0.2760	0.2872	0.3449	0.3034	0.4434	0.3245
PME	0.1477	0.1250	0.1340	0.1150	0.2162	0.2744	0.1687
PPP	2.5674	1.7048	1.7184	1.5905	4.0381	2.7170	2.3894
SPM	1.0000	1.0000	1.0000	0.8724	0.3005	1.2068	0.8966
TRA	0.1922	0.2515	0.2254	0.1388	0.1954	0.2105	0.2023
VDP	0.6272	0.8359	0.5231	0.4758	0.5877	0.7144	0.6273

Table 5. Difference between business performance

DMUs	Q1/2019	Q2/2019	Q3/2019	Q4/2019	Q1/2020	Q2/2020	Average
DBD	0.7082	0.7914	0.7652	1.1616	1.0360	0.7465	0.8681
DCL	-7.9143	-3.4968	0.1793	1.8179	0.0229	0.1800	-1.5352
DHG	0.9197	1.0167	0.9538	1.1242	1.1160	1.1110	1.0402
DMC	1.0059	1.4895	1.3280	1.3092	0.8486	1.4659	1.2412
DP3	0.5711	2.7401	0.3779	-0.5996	-0.0928	0.3232	0.5533
IMP	0.4165	0.4459	0.5247	0.9512	0.5045	0.6540	0.5828
MKV	0.7490	0.8551	0.8660	0.6028	0.8095	0.4070	0.7149
OPC	0.6924	0.8842	0.7546	1.9475	0.8145	0.6342	0.9546
PME	1.2755	1.4905	2.4966	2.3873	-0.3943	0.7380	1.3323
PPP	-0.4085	-0.5096	-0.3497	-0.2007	0.1558	-0.6848	-0.3329
TRA	0.9327	1.1076	0.7323	1.3694	0.8982	1.1482	1.0314
VDP	0.5935	0.3805	0.6480	0.7216	0.5487	0.2327	0.5208

Assigning values to the input and output variables with the presence of negative values such as $x_i^{\min}, y_i^{\min}$ $(i = 1, K, m)$, respectively. MI is calculated as follows:

$$\min_{\theta,\lambda} \theta = D^{\min}(x_i^{\min}, y_i^{\min}) \tag{4}$$

In which

$$\begin{aligned}
&x_i^{\min} = Min(x_{i1}^1, x_{i2}^1, K, x_{i1}^T, K, x_{in}^T) \\
&i = 1, K, m) \\
&x_i^{\min} \geq 0, x_i^{\min} = 0 / x_i^{\min} < 0, x_i^{\min} \times 1.1 \\
&y_i^{\min} = Min(y_{i1}^1, y_{i2}^1, K, y_{i1}^T, K, y_{in}^T) \\
&i = 1, K, m \\
&y_i^{\min} \geq 0, y_i^{\min} = 0 / y_i^{\min} < 0, y_i^{\min} \times 1.1
\end{aligned} \tag{5}$$

Narjes and Alireza (2014) showed that results indicating an increase in efficiency are shown as greater than 1. On the other hand, results whose values are less than 1 imply a marked reduction in efficiency, these inefficient companies should have a solution to increase the efficiency.

3. Results and discussions

In the study, data from the 12 pharmaceutical companies listed on the stock market in Vietnam since early 2019 to the end of Q2/2020 was used in the analysis. Table 2 summarizes the variables used in efficiency analysis.

According to Table 1, the statistical data revealed the presence of negative values including PAT of Q1 and Q2 in 2019 and EPS of Q2/2019. As the normal models in DEA, these values must remove, however, based on the characteristics of the Malmquist Model which allows for negative values, the annual efficiency of each pharmaceutical company was still able to be calculated. Before going into the data calculation method, the initial values need to be checked for correlation to ensure a relationship between the variables according to the principles in DEA.

Correlative relationships between variables must always range between -1 and +1. Based on the results of the analysis, the correlation coefficient between the variables in stage one range between 0.7696 and 1, indicating a relatively good correlation in the business performance stage. In the second stage of market return efficiency, the variables are correlated from – 0.2225 to 1. Despite not indicating good results, these correlation values are still within an acceptable range to indicate efficiency. Therefore, all the input and output data in the two analysis stages are suitable for an application to the Malmquist model wherever negative values appear.

For any unit operating in business or production, the first stage of enterprise always involves capital investment, followed by profits with interest from sales after a period of sustained operations. To provide concrete results regarding the business efficiency of the pharmaceutical industry in Vietnam, the study used both input and output variables by applying the technical model to calculate the scores, as shown in Table 3.

The business performance results described 12 pharmaceutical companies in Vietnam, as shown in Table 1, showed considerable fluctuations across different operating periods. Analysis revealed that three of the companies: DHG, DMC, and PPP always maintained business efficiency from early 2019, where the scores in each period were above 1 and the average value of all six quarters for each of the companies is 1.1146; 1,5104 and 3.7216, respectively. SPM, meanwhile, had the lowest performance for all six quarters, with scores below 1 and an average score of 0.5637. Other companies have fluctuating scores in each quarter, which indicated both efficient and inefficient periods. DCL and IMP, with average scores below 1, did not achieve the efficiency score, possessing the results of 0.8210 and 0.7955, respectively. During the same period, the remaining pharmaceutical companies had ineffective periods, but their average scores were still above 1. The results, thus, implied that the SPM pharmaceutical company has the worst comparative business performance.

Based on the scores of these pharmaceutical companies in the business activities, their positions in this stage were determined as shown in Figure 2. PPP was considered as the best company when it always held the first position in 5 periods except for 2019/Q2. DP3 tried to best to rank the first position in Q2/2019, but it has been down sharply in Q4/2019, and then kept stable within two earlier quarters of 2020. DMC only ranked the second position in Q3/2019, it owns the third and fourth positions with four periods and one period, respectively. PME had a unique period with the second position, remaining periods ranked the eighth and ninth position. DBD, DHG, OPC, TRA, and VDP ranked in the middle position from fourth to tenth. DCL only ranked the third position in Q4/2019, other periods had the eleventh and twelfth positions. SPM was the worst company that always stayed at the eleventh and twelfth position from Q1/2019 to Q2/2020.

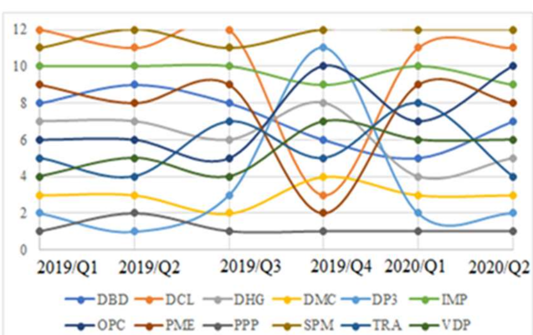


Figure 2. Ranking the business activities

In stage 1, efficient and inefficient scores and the ranking of each pharmaceutical company in each quarter from the beginning of 2019 to the end of Q2/2020 regarding business performance were determined. In stage 2, market efficiency was based on the input variables taken from the output variables in stage one and based on the output index. The paper calculated market efficiency based on the Malmquist model, as shown in Table 4.

The results of market efficiency research showed that there are only two pharmaceutical companies, DP3 and PPP, that achieved effective results during all six quarters examined, and they had average scores of over 1, with 1.3498 and 2.3894, respectively. In contrast, up to eight pharmaceutical companies, including DBD, DHG, DMC, IMP, OPC, PME, TRA, and VDP, did not have any efficient operational periods and their average scores were all lower than a value of one, their average scores were 0.1786; 0.0744; 0.2689; 0.2127; 0.3245; 0.1687, 0.2023, and 0.6273, respectively. The other two companies, DCL and SPM, have both efficient and inefficient periods. Whereas SPM had an efficient average score of only 0.8966, and two inefficient and four efficient periods; although DCL experienced a period of inefficiency, the average score is still above 1 (2.3561), it had two efficient and four inefficient periods. Based on the average score, PME was considered as the worst company with the lowest average score (0.1687), and PPP was determined the best company with the highest average score as 2.3894.

The results of market efficiency research showed that there are only two pharmaceutical companies, DP3 and PPP, that achieved effective results during all six quarters examined, and they had average scores of over 1, with 1.3498 and 2.3894, respectively. In contrast, up to eight pharmaceutical companies, including DBD, DHG, DMC, IMP, OPC, PME, TRA, and VDP, did not have any efficient operational periods and their average scores were all lower than a value of one, their average scores were 0.1786; 0.0744; 0.2689; 0.2127; 0.3245; 0.1687, 0.2023, and 0.6273, respectively. The other two companies, DCL and SPM, have both efficient and inefficient periods. Whereas SPM had an efficient average score of only 0.8966, and two inefficient and four efficient periods; although DCL experienced a period of inefficiency, the average score is still above 1 (2.3561), it had two efficient and four inefficient periods. Based on the average score, PME was considered as the worst company with the lowest average score (0.1687), and PPP was determined the best company with the highest average score as 2.3894.

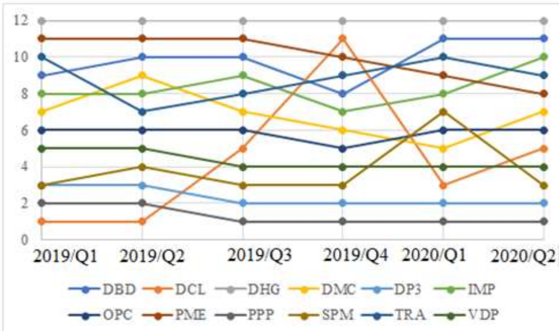


Figure 3. Ranking the market activities

As same as the first stage, the position of each pharmaceutical company in the market activities was determined based on the score of the market activities in Table 4. Figure 3 described the continual variation of 12 pharmaceutical companies in each period. PPP ranked the second position in Q1 and Q2 of 2019, and then it tried the best to hold the first position in the remaining periods. DP3 was similar to PPP when this company owned the third position in Q1 and Q2 of 2019 and the second position in the remaining periods. DCL had excellent quarters with the first position in Q1 and Q2 of 2019; however, it was down sharply, especially in Q4/2019 with the eleventh position. DHG was defined as the worst company when it always held the final position the whole time. The remaining companies had consecutive variations in each period.

There are clear differences between the business performance and market performance of each pharmaceutical company during each quarter. Based on the business performance scores provided in Table 3, as well as the market performance scores in Table 4, the difference between the two stages was calculated using formula (4). The detailed results are given in Table 5.

$$D = HQKD - HQTT \tag{6}$$

Where D is the difference score; HQKD is business efficiency; HQTT is the market efficiency.

Table 5 showed that most pharmaceutical companies have higher business performance scores than market performance scores, except for a few cases with higher market performance scores, DCL in Q1 and Q2 of 2019; DP3 in Q4 of 2019 and Q1 of 2020; PME in Q1 of 2020; PPP in most quarters except Q1 of 2020. Average scores for DCL and PPP, whose business efficiency scores are lower than market efficiency, are -1.5352, and -0.3329, respectively. According to the analysis, only PPP achieved effective scores throughout the entire period in both stages of business efficiency and market performance with average scores for each period all above a value of 1.

Based on the empirical results shown and DEA, one suggestion is to increase the score during inefficient periods, to increase output variables, and to decrease input variables. In the first stage, input values for inefficient units need to reduce excess in input variables, such as total assets, operating costs, and equity. Obviously, further effort needs to be made to achieve better business results. In the second stage, during relative periods of market inefficiency, companies need to exhibit an effective strategy to increase earnings per share and market value per share.

4. Conclusion

Input and output variables of pharmaceutical companies related to business performance and market performance were identified and analyzed in this research. The paper studied twelve pharmaceutical companies listed on the Vietnam stock market during the period from early 2019 to the end of Q2/2020. Results provide a detailed account of efficiency in business activities and business marketing for each pharmaceutical company, in every quarter, thus determining efficient and inefficient operational periods. Final results reveal, that of all twelve pharmaceutical companies, MKV achieved the best efficiency throughout the whole assessed period, thus, obtaining optimum efficiency in the marketplace.

Research results also show that many companies experienced periods of ineffective business and diminished market efficiency. These values offered insights into the inefficient operational period of each company. It is, therefore, advisable that a reassessment takes place of the input and output variables. This must be done to better strategize and to find a suitable solution for improving future efficiency levels. If during the business performance evaluation process, there are adjustments to reduce input variables and increase output variables, the value will be improved significantly. For the market assessment process, the output variables need to follow a strategy and deliberate policies to improve.

Although the study indicated the performance of pharmaceutical companies in Vietnam for both business and market processes, it still has limitations. First, the study only demanded the pharmaceutical industry in Vietnam, the further research should compare with the pharmaceutical industry of other countries. Second, the research measured the efficiency of 12 pharmaceutical companies in the previous time, the further study can utilize the forecasting model such as GM (1,1) model, Holt-winters methods to have a foreseen observation.

Reference

Aghayi, N., Tavana, M., & Maleki, B. (2019). *A Malmquist productivity index with the directional distance function and uncertain data*. Scientia Iranica, 26(6), 3819-3834. doi: 10.24200/sci.2018.5259.1173.

An, Q. X., Wu, Q. F., Li, J. L., Xiong, B. B., & Chen, X. H. (2019). *Environmental efficiency evaluation for Xiangjiang River basin cities based on an improved SBM model and Global Malmquist index.* Energy Economics, 81, 95–103. Available at: https://doi.org/10.1016/j.eneco.2019.03.022

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). *The economic theory of index numbers and the measurement of input, output and productivity.* Econometrica, 50(6), 1393–1414. Available at: https://doi.org/10.2307/1913388

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). *Measuring the efficiency of decision-making units*. European Journal of Operational Research, 2(6), 429–444. Available at: https://doi.org/10.1016/0377-2217(78)90138-8

David, T. (2015). *The Pharmaceutical Industry and the Future of Drug Development*. Pharmaceuticals in the Environment, 1-33 DOI: 10.1039/9781782622345-00001

Elmhirst, E., & Urquhart, L. (2020). *Big pharma sports smallest late-stage pipeline for a decade.* Evaluate Vantage, 1–20. Available at: https://www.evaluate.com/vantage/articles/data-insights/ma/big-pharma-sports-smallest-late-stage-pipeline-decade

Esmaeili, F. S., & Malkhalifeh, M. R. (2017). *Data Envelopment Analysis with Fixed Inputs, Undesirable Outputs and Negative Data*. Data Envelopment Analysis and Decision Science, 1, 1–6. DOI: 10.5899/2017/dea-00140

Fare, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). *Productivity growth, technical progress, and efficiency change in industrialized countries*. The American Economic Review, 84(1), 66–83. Available at: https://www.jstor.org/stable/2117971

Ghasem, T., Shabnam, R., & Simin, T. (2014). *Profit Malmquist Index and Its Global Form in the Presence of the Negative Data in DEA*. Journal of Applied Mathematics, 2014, Article ID 276092, 1–8. Available at: https://doi.org/10.1155/2014/276092

Leung, A. Y. (2006). *Traditional toxicity documentation of Chinese Materia Medica--an overview*. Toxicologic Pathology, 34(4), 319-26. Available at: https://doi.org/10.1080/01926230600773958

Maria, C. A. S. P., & Emmanuel, T. (2010). *Malmquist-type indices in the presence of negative data: An application to bank branches*. Journal of Banking & Finance, 34(7), 1472–1483. Available at: https://doi.org/10.1016/j.jbankfin.2010.01.004

Narjes, M., & Alireza, Y. (2014). *The Biennial Malmquist Index in the of Negative Data*. Journal of Mathematics and Computer Science, 12(1), 1–11. Available at: http://dx.doi.org/10.22436/jmcs.012.01.01

Naser, A., Hadi, B. V., & Ali, E. (2018). *Application of Malmquist productivity index with carry-overs in power industry.* Alexandria Engineering Journal, 57(4), 3151–3165. Available at: https://doi.org/10.1016/j.aej.2017.11.007

Seiford L. M., & Zhu, J. (2002). *Modeling undesirable factors in efficiency evaluation.* European Journal of Operational Research, 142(1), 16–20. Available at: https://doi.org/10.1016/S0377-2217(01)00293-4

Tone, K. (2002). *A slacks-based measure of super-efficiency in data envelopment analysis*. European Journal of Operational Research, 143 (1), 32–41. Available at: https://doi.org/10.1016/S0377-2217(01)00324-1

Tone, K. (2016). *Malmquist Productivity Index*. Handbook on Data Envelopment Analysis, 203–227. Doi: 10.1007/1-4020-7798-X_8.

Vietstock. Available online: https://vietstock.vn/. Accessed on 2 October 2020

OPTIMIZING PRODUCTION FLOW BY IMPLEMENTATION GROUP TECHNOLOGY: A CASE STUDY IN EDUCATIONAL EQUIPMENT SHOP FLOOR

Truong Thanh Tam[1], Nguyen Hoang Phuc[2], Do Ngoc Hien[3]

[1] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh city University of Technology – Ho Chi Minh National University, tttam.sdh20@hcmut.edu.vn*

[2] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh city University of Technology – Ho Chi Minh National University, nhphuc.sdh20@hcmut.edu.vn*

[3] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh city University of Technology – Ho Chi Minh National University, hienise97@hcmut.edu.vn*

Abstract: Layout design is one of the most important techniques in lean manufacturing. Nowadays, layout design before mass production has been of interest to all companies in the manufacturing industries because it contributed many benefits to the company. Especially, in the mechanical manufacturing industry with the batch production, the reason comes from limited resources such as machines and equipment as well as people. This paper investigates the effectiveness of layout design in terms of cellular manufacturing approach with the aim of optimizing the flow of product components to increase factory efficiency. The current layout will be identified by using the Group technology method (GT). With the proposed method, the existing layout was observed and analyzed workflow step by step of each product. Through it, machines that handle the product component have the similar properties will be grouped together. It allows the production process for many different parts in small volumes to still achieve the efficiency without product standardization. According to the results after the redesign, the production process will be improved, the transportation time between machines in the process will be significantly reduced, and the process flow will be smoother. It shows that the GT method is effective for analyzing workflow to solve productivity problems.

Keywords: Cellular manufacturing; Group technology; Lean manufacturing.

1. Introduction

Cellular manufacturing is one of the forms of JIT manufacturing and lean manufacturing with the aim of minimizing transportation time, producing many similar products at the same time with minimal waste. It was proposed by Flanders in 1925 and promoted by Burbidge (1975). According to this method, each cell consists of one or more different machines performing a certain task. The product moves from one cell to the next cell, each workstation will complete a part of the production process (Liker and Jeffery, 2004). Besides, Group technology (GT) based on small batches production of different parts to achieve the aim of economics if they are grouped and scheduled for production according to characteristics. Group technology makes most of the similarities about product components by using the processes, tools, and machines to manufacture them. Accordingly, the machines are grouped into one cell, each cell being responsible for producing a family of parts (Ivanovo E.K, 1968).

One of the advantages of cellular manufacturing is its flexibility. Because most changes are done quickly on automated machines. This allows scaling up production for a product with small changes to the overall design or a complete change to the overall design (Liker and Jeffery, 2004). In order to apply cellular production, several steps must be considered and performed, such as: the similar parts must be grouped together in the first step, then a workflow analysis of each group must be performed (PFA) (Inman et al, 2006). Currently, there are many optimization algorithms for group machines and part was proposed as Rank Order Cluster, Modified Order Cluster (Amruthnath and Nagdev; Gupta and Tarun 2016). There are also a number of mathematical models that support planning for a cell production center in a multi-target environment such as multiple plant locations, multi-market allocations with production planning and various part mix (Aalaei et al, 2017).

With many benefits from the GT method bring out for any company, there are many research in this field but most of them are methodological studies such as (Tamal Ghosh et al, 2016) and (Wafik Hachicha, 2007) while the number of experimental studies to solve the problem in actual production is very low as well. one of the related researches is the one by Nikola Suzić (2012), the paper proposed GT method in the furniture manufacturing company. With 440 different products, Group technology method was used for 16 machines in the workshop to optimize the materials flow. In their article, products have similarities in characteristics classified as grouped together. Through the from/to chart, the work flow is determined to analyze the operation process and the new layout was created. As a result, setting up time was reduced 3 times, the raw materials flow was simpler and shortened lead time was achieved specially.

The purpose of this research is to highlight the effectiveness of cell manufacturing oriented layout redesign using GT techniques to group devices by cell in a job shop environment. It also shows that the production flow will be improved, and the transit time between many machines in the process will be significantly reduced. From there, the lead time of process production is reduced through the waste of transportation time and waiting time were identified and eliminated.

2. Literature Review

2.1. Cellular Manufacturing

Cell manufacturing is an important part of lean manufacturing and JIT includes group technology. The goal of cell manufacturing is to move the semi-finished product as quickly as possible, creating more similar products while creating less waste (Liker, Jeffery, 2004). Cellular production was proposed by Flanders in 1925 and applied by Mitrofanov in Russia in 1933. Clustering creates a system that utilizes similarities in components, processes or products to provide benefits including reduced up time, reduced work flow, reduced inventory and material handling costs, and improved product quality (N Hyer et al, 2002). The proposed cell layout model was shown in **Figure 1** as below

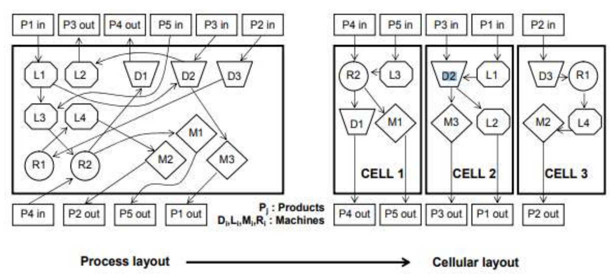


Figure 1. Cellular layout model, N.Hyer et al (2002)

Cellular is created by aggregating activities or machines or people involved in a production process and grouping them together and separate from another one. Cellular combines the processes required to create a particular product, which minimize unrelated steps in the process while at the same time facilitating determination to quickly identify problems and encourage staff in the cell to solve emerging problems quickly (Morgan, J.M, 2006).

2.2. Euclidean Distance Matrix (EDM)

In a multidimensional space, the sum of squared deviations in their coordinates is called the Euclidean distance. Euclidean distance matrix analysis was developed in the 1990s, however the widespread use of these matrices in growth analysis was developed by Richtsmeier and Lele (1993), it is considered the distance estimation tools, there are important statistical properties, namely the consistency and asymptotic attributes. In mathematics, the Euclidean distance matrix is a square matrix with n elements. It represents the distance of a set of n points $\{x_i, i = 1, . . . , n\}$ in the Euclidean space R^n .EDM is a square matrix and the measure of distance-square between points of xi and $x j$, is calculated as:

$$A = (a_{ij}) \text{ where } a_{ij} = d_{ij}^2 = (x_i - x_j)^2 \qquad (1)$$

The EDM matrix model was shown in **Figure 2** as below:

$$A = \begin{bmatrix} 0 & d_{12}^2 & d_{13}^2 & \cdots & d_{1n}^2 \\ d_{21}^2 & 0 & d_{23}^2 & \cdots & d_{2n}^2 \\ d_{31}^2 & d_{32}^2 & 0 & \cdots & d_{3n}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1}^2 & d_{n2}^2 & d_{n3}^2 & \cdots & 0 \end{bmatrix}$$

Figure 2. EDM matrix by Dokmanic et al (2015)

The Properties of EDM matrix:

- All parts on the diagonal of the EDM matrix are "0" number.
- A is symmetrical, which means that the values on both sides of the diagonal are the same ($a_{ij} = a_{ji}$)
- A is property of the triangle inequality :
$$\sqrt{a_{ij}} \le \sqrt{a_{ik}} + \sqrt{a_{kj}}$$
- $a_{ij} \ge 0$

2.3. Non-hierarchical clustering method

A heuristic approach based on Euclidean distance matrix was introduced that consisted of seven main steps (Manash Hazarika, 2016) which were suggested as follows:

Step 1. Convert the machine-part into binary matrix

Convert the machine-component data into "m × p" binary incidence matrix (with **m** (columns) is machine quantity and **p** (rows) is the number of component):

$$a_{ij} = \begin{cases} 1 & \text{, if machine i process the part j} \\ 0 & \text{otherwise} \end{cases}$$

Step 2. Convert the binary matrix into machine-part volume matrix

To convert a binary matrix into a machine part mass matrix, multiply the binary matrix by volume's parts matrix (matrix V)

V = [v_1, v_2….v_j], (where: v_j is parts volume)

Step 3. Standardized the machine-part volume incidence matrix

The goal of this step is eliminate data bias, return them to normalized form using the formula involving standard deviation:

- To standardize the matrix B, calculate the sum of each individually column in the first times:
$$B = \sum_{j=1}^{p} b_{ij} \quad (2)$$
Where: i = 1, 2…m and j = 1, 2…p
- Secondly, we calculate average of B:
$$\overline{B_i} = \frac{B_i}{p} \quad (3)$$
- Thirdly, finding the variance of B:
$$\partial_i^{\,2} = \overline{B_i} - (\overline{B_i})^2 \quad (4)$$
- Finally, The standardized matrix (matrix C) was calculated:
$$c_{ij} = \frac{(b_{ij} - \overline{B_i})}{\partial_i} \quad (5)$$

Step 4. Calculate the Euclidean distance matrix (EDM):

The Euclidean distance between machines **x** and **y** is defined as follows:

$$d_{xy} = \sqrt{\sum_{j=1}^{p} (c_{jx} - c_{jy})^2} \quad (6)$$

Respectively calculated for the remaining machine we obtain the EDM matrix:

$$A = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1m} \\ d_{21} & d_{22} & \dots & d_{2m} \\ \dots & \dots & \dots & \dots \\ d_{m1} & d_{m2} & \dots & d_{mm} \end{bmatrix}$$

Step 5. Group machines according to the required cell count.

Two machines with the smallest Euclidean distance value are grouped with together until the required number of machines in one cell is over or the number of cell is over

Step 6. Group special machines.

The special machines are defined as cannot be moved, limited in quantity or handle multiple products at the same time. To merge these machines into the cell, the volume value at each expected cell was calculated, special machines will be merged at the cell with the largest volume value.

Step 7. Cluster the parts for part families up to the number of cells.

For similar product families, when there is not enough information on quantity, we can choose the best minimum distance route between cells or maximum movement within one cell.

2.4. Group Technology efficiency

To evaluate the quality of a cell formation solution, an evaluation index was proposed by Kumar and Chandrasekharan (1990) is GE (Group efficiency). It was defined as the proportion of machines arranged in a cell and out of a cell, as determined by a formula as below:

$$\tau = \frac{n_1 - n_1^{out}}{n_1 + n_0^{in}} = \frac{n_1^{in}}{n_1 + n_0^{in}} \quad (7)$$

Where:

- n_1 : The number of "1" in the final machine-product matrix
- n_1^{out} : The number of "1" out of the cell
- n_1^{in} : The number of the number "1" inside the cell
- n_0^{in} : The number of the number "0" inside the cell

This formula has been shown to better reflect the quality of a cell formation solution

3. Optimizing production flow: a case study

3.1. Current layout

A typical research case is an educational equipment factory in Vietnam. The research is applied to a traditional mechanical workshop with many specialized machines where the machine's efficiency is almost based on the operation's skill. Existing the layout was arranged according to functions where each area was a group of different machines to produce a variety of products. An AutoCAD drawing of the current layout shows the positions of all machines and the workflow of products as below:

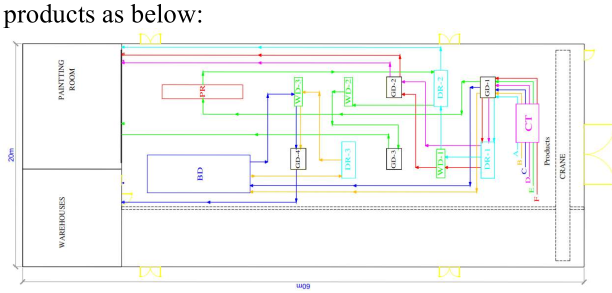


Figure 3. An AutoCAD drawing current layout

In this process, the cutting and the painting area are arranged separately and fixed, so this research focuses on 12 machines left from Grind 1 to Weld 3 only for analysis and re-layout. For the convenience of the calculation, the machines are encoded in the order from 1 to 12 as table below:

Table 1. Machine number

Machine	Number
Grind 1	1
Grind 2	2
Grind 3	3
Grind 4	4
Drill 1	5
Drill 2	6
Drill 3	7
Press	8
Bend	9
Weld 1	10
Weld 2	11
Weld 3	12

The sequence of process information and production volume to produce 6 product's type was shown in **Table 2** below:

Table 2. Machine sequence information all products

No.	Products	Process	Volume (Pcs/ month)
A	Computer Table	1 - 5 - 10 - 2	2500
B	Bed for Dormitory	1 - 6 - 11 - 3	500
C	Shoe Rack	1 - 8 - 6 - 11 - 3	800
D	Board	1 - 5 - 2	1000
E	Book Shelves	1 - 9 - 12 - 4	800
F	Chair	1 - 9 - 7 - 12 - 4	5000

3.2. Implementation of proposed approach

In this paper, an algorithmic approach based on the Euclidean Distance matrix was proposed for solving the current layout problem. It was done through 7 specific steps as follows:

Step 1. Convert the machine-part into binary matrix

Firstly, the product-machine matrix is built that reflects the relationship between products and machine as picture below:

Products	Procduct Volume	Machine											
		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
A	2500	1	4			2					3		
B	500	1		4			2					3	
C	800	1		5			3		2			4	
D	1000	1	3			2							
E	800	1			4					2			3
F	5000	1			5			3		2			4

Figure 4. Product – machine matrix

Secondly, the machine matrix is converted into a binary matrix (6 × 12) with 6 products and 12 machines through the principle:

$$a_{ij} = \begin{cases} 1, \text{if machine i process the part j} \\ 0, \text{if otherwise} \end{cases}$$

Where each row represents a specific product and each column represents a specific machine.

Step 2. Convert the binary matrix into machine-part volume matrix

After the product-machine matrix was built, the machine-product volume matrix was created by multiplying the machine product matrix and the product volume matrix by the formula:

B =A.V

With: V is the products volume matrix:

$$V = \begin{bmatrix} 2500 \\ 500 \\ 800 \\ 1000 \\ 800 \\ 5000 \end{bmatrix}$$

And finally, machine-product volume incidence matrix was calculated as below:

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} 2500 \\ 500 \\ 800 \\ 1000 \\ 800 \\ 5000 \end{bmatrix}$$

After multiplied the machine product matrix and the product volume matrix, the result was shown as below:

$$B = \begin{bmatrix} 2500 & 2500 & 0 & 0 & 2500 & 0 & 0 & 0 & 0 & 2500 & 0 & 0 \\ 500 & 0 & 500 & 0 & 0 & 500 & 0 & 0 & 0 & 0 & 500 & 0 \\ 800 & 0 & 800 & 0 & 0 & 800 & 0 & 800 & 0 & 0 & 800 & 0 \\ 1000 & 1000 & 0 & 0 & 1000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 800 & 0 & 0 & 800 & 0 & 0 & 0 & 0 & 800 & 0 & 0 & 800 \\ 5000 & 0 & 0 & 5000 & 0 & 0 & 1 & 0 & 5000 & 0 & 0 & 1 \end{bmatrix}$$

Step 3. Standardized the machine-part volume incidence matrix

Firstly, to standardize the machine-part volume incidence matrix (matrix B), the sum of each column individually was calculated by the formula number (2), as below:

$$B_1 = \sum_{j=1}^{p} bj1 = 10600$$

$$B_2 = \sum_{j=1}^{p} bj2 = 3500$$

…..

$$B_{12} = \sum_{j=1}^{p} bj8 = 5800$$

Secondly, an average of B_i was calculated by using formula number (3) as follows:

$$\overline{B}_1 = \frac{B_1}{P} = \frac{10600}{6} = 1766.67$$

$$\overline{B}_2 = \frac{B_2}{P} = \frac{3500}{6} = 583.33$$

…..

$$\overline{B}_{12} = \frac{B_8}{P} = \frac{3500}{6} = 966.67$$

Thirdly, the standard deviation (∂_i) was calculated based on the formula (4) as follows:

$$\partial_1 = \sqrt{|\overline{B}_1 - (\overline{B}_1)^2|} = 1766.17$$

$$\partial_2 = \sqrt{|\overline{B_2} - (\overline{B}_2)^2|} = 582.83$$

…..

$$\partial_{12} = \sqrt{|\overline{B}_{12} - (\overline{B}_{12})^2|} = 966.17$$

Finally, applying formula number (5), we have:

$$C_{11} = \frac{(2500 - 1766.67)}{1766.17} = 0.42$$

$$C_{21} = \frac{(500 - 1766.67)}{1766.17} = -0.72$$

…..

$$C_{81} = \frac{(10600 - 1766.67)}{1766.17} = -1.00$$

With the similar calculation, the standardized machine-product volume incidence matrix was shown in below:

$$C = \begin{bmatrix} 0.42 & 3.29 & -1 & -1 & 3.29 & -1 & -1 & -1 & -1 & 5 & -1 & -1 \\ -0.72 & -1 & 1.31 & -1 & 4 & 1.31 & -1 & -1 & -1 & 4 & 1.31 & -1 \\ -0.55 & -1 & 2.7 & -1 & 4 & 2.7 & -1 & 5.02 & -1 & 4 & 2.7 & -1 \\ -0.43 & -0.71 & -1 & -1 & 0.71 & -1 & -1 & -1 & -1 & 4 & -1 & -1 \\ -0.55 & -1 & -1 & -0.17 & 4 & -1 & -1 & -1 & -0.17 & 4 & -1 & -0.17 \\ 1.83 & -1 & -1 & 4.17 & 4 & -1 & 5 & -1 & 4.17 & 4 & -1 & 4.17 \end{bmatrix}$$

Step 4. Calculate the Euclidean distance matrix (EDM):

The Euclidean distance between 2 machines is determined by formula (6). Firstly, we calculate for machine 01 and machine 02 as below:

$$d_{12} = d_{21} = \sqrt{\sum_{j=1}^{p}(c_{j1} - c_{j2})^2} = 4.25$$

$$d_{13} = d_{31} = \sqrt{\sum_{j=1}^{p}(c_{j1} - c_{j3})^2} = 5.02$$

With the similar calculation, the Euclidean distance matrix (D) was shown in Figure 5 as below:

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
M1	0.00	4.25	5.02	2.87	4.25	5.02	3.59	6.45	2.87	5.47	5.02	2.87
M2	4.25	0.00	6.36	6.99	0.00	6.36	7.58	7.59	6.99	2.43	6.36	6.99
M3	5.02	6.36	0.00	6.82	6.36	0.00	7.42	3.28	6.82	7.43	0.00	6.82
M4	2.87	6.99	6.82	0.00	6.99	6.82	1.17	7.98	0.00	7.97	6.82	0.00
M5	4.25	0.00	6.36	6.99	0.00	6.36	7.58	7.59	6.99	2.43	6.36	6.99
M6	5.02	6.36	0.00	6.82	6.36	0.00	7.42	3.28	6.82	7.43	0.00	6.82
M7	3.59	7.58	7.42	1.17	7.58	7.42	0.00	8.50	1.17	8.49	7.42	1.17
M8	6.45	7.59	3.28	7.98	7.59	3.28	8.50	0.00	7.98	8.51	3.28	7.98
M9	2.87	6.99	6.82	0.00	6.99	6.82	1.17	7.98	0.00	7.97	6.82	0.00
M10	5.47	2.43	7.43	7.97	2.43	7.43	8.49	8.51	7.97	0.00	7.43	7.97
M11	5.02	6.36	0.00	6.82	6.36	0.00	7.42	3.28	6.82	7.43	0.00	6.82
M12	2.87	6.99	6.82	0.00	6.99	6.82	1.17	7.98	0.00	7.97	6.82	0.00

Figure 5. Euclidean distance matrix (D) for the case study

Due to the symmetry of the Euclidean matrix, the two sides of the diagonal are similar to each other ($d_{12}=d_{21}$). We can reduce the Euclidean distance matrix (D) as follows:

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
M1	0.00											
M2	4.25	0.00										
M3	5.02	6.36	0.00									
M4	2.87	6.99	6.82	0.00								
M5	4.25	0.00	6.36	6.99	0.00							
M6	5.02	6.36	0.00	6.82	6.36	0.00						
M7	3.59	7.58	7.42	1.17	7.58	7.42	0.00					
M8	6.45	7.59	3.28	7.98	7.59	3.28	8.50	0.00				
M9	2.87	6.99	6.82	**0.00**	6.99	6.82	1.17	7.98	0.00			
M10	5.47	2.43	7.43	7.97	2.43	7.43	8.49	8.51	7.97	0.00		
M11	5.02	6.36	0.00	6.82	6.36	**0.00**	7.42	3.28	6.82	7.43	0.00	
M12	2.87	6.99	6.82	0.00	6.99	6.82	1.17	7.98	**0.00**	7.97	6.82	0.00

Figure 6. Euclidean distance matrix (D) after reduced

Step 5. Cluster the machines up to required number of cells

According to the Euclidean distance matrix as shown in **Figure 5**, we have the smallest EDM distance between M4-M9, M6-M11 and M9-M12, so we merge machine group 4-9-12 and machine group 6-11 into two entities unique, and then consider the smallest value of the Euclidean distance from these two groups of entities to the other machines that was shown in Figure 7 below:

	M1	M2	M3	M4-M9-M12	M5	M6-M11	M7	M8	M10
M1	0.00								
M2	4.25	0.00							
M3	5.02	6.36	0.00						
M4-M9-M12	2.87	6.99	6.82	0.00					
M5	4.25	0.00	6.36	6.99	0.00				
M6-M11	5.02	6.36	0.00	6.82	6.36	0.00			
M7	3.59	7.58	7.42	1.17	7.58	7.42	0.00		
M8	6.45	7.59	3.28	7.98	7.59	3.28	8.50	0.00	
M10	5.47	2.43	7.43	7.97	2.43	7.43	8.49	8.51	0.00

Figure 7. Euclidean distance matrix (D) after merge machine group 4-9-12 and machine group 6-11

The analysis results after merging two groups of machines (6-11 and 4-9-12) show that the value of the Euclidean distance of the machines has changed, especially the Euclidean distance of M2 and M5 is the smallest and the Euclidean distance of M2 and M5 is the smallest, machine (6 -11) and M3 are also minimal. So the next step is to merge the M2-M5 machine group and the M6-M11-M3 machine group together as shown in Figure 8 below:

	M1	M2-M5	M4-M9-M12	M6-M11-M3	M7	M8	M10
M1	0.00						
M2-M5	4.25	0.00					
M4-M9-M12	2.87	6.99	0.00				
M6-M11-M3	5.02	6.36	6.82	0.00			
M7	3.59	7.58	1.17	7.42	0.00		
M8	6.45	7.59	7.98	3.28	8.50	0.00	
M10	5.47	2.43	7.97	7.43	8.49	8.51	0.00

Figure 8. Euclidean distance matrix (D) after merge machine group M2-M5 and machine group 6-11-13

Performing with the remaining machines in turn based on the smallest Euclidean distance, we get the result of the Euclidean distance matrix after grouping all the machines as shown in Figure 9:

	M2-M5-M10	M4-M9-M12-M7-M1	M6-M11-M3-M8
M2-M5-M10	0.00		
M4-M9-M12-M7-M1	6.99	0.00	
M6-M11-M3-M8	6.36	5.02	0.00

Figure 9. The Euclidean Distance Matrix after clustering all machines

According to the above results, three machine cells are formed as follows:

- Cell 1: includes machines M2-M5-M10
- Cell 2: includes machines M4-M9-M12-M7-M1
- Cell 3: includes machines M6-M11-M3-M8

Step 6. Group special machines.

In this case, **MC1** is designated as a special machine because it is a limited quantity machine as well as handles most of the stages of all products, so it is arranged into a cell with a total largest product family volume (E and F).

Step 7. Group the product families corresponding to the calculated cell number

By selecting the best routes of products with minimal intercellular movement, the final product families selected will be:

- Product family 01: A and D
- Product family 02: E and F
- Product family 03: B and C

The final cell formation and product family are shown in the Figure 10:

Products	Procduct Volume	Machine											
		M2	M10	M5	M4	M7	M12	M9	M1	M8	M6	M11	M3
A	2500	4	3	2					1				
D	1000	3	0	2					1				
E	800				4	0	3	2	1				
F	5000				5	3	4	2	1				
B	500								1	0	2	3	4
C	800								1	2	3	4	5

Figure 10. The final solution

To evaluate the efficiency when grouping machines into each cell, a proposed evaluation formula is Group Technology efficiency as formula number (6) was presented in part 2:

$$\tau = \frac{n_1 - {n_1}^{out}}{n_1 + {n_0}^{in}} = \frac{25-4}{25+3} = 75\%$$

After analysis and calculation, the machines will be arranged into the cells, the results are presented as shown in the following figure:

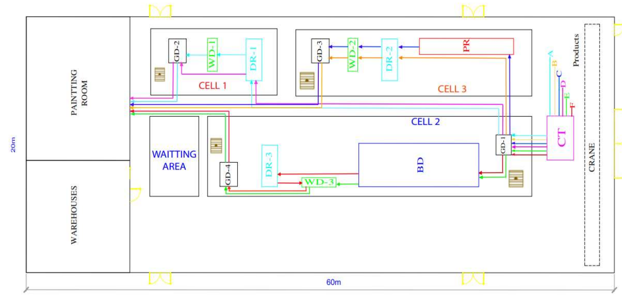


Figure 11. Cellular arrangement

4. Results and discussion

The efficiency of the factory depends not only on the skills of the workers but also on the efficient arrangement of machines and resources. According to the analysis results, the workflow was really "smooth" compared to the previous layout. Through the application of the proposed method, the Workflow and efficiency of GT are almost optimized (can achieve 75%).

To demonstrate the effectiveness of the method, a comparison model with two widely applied algorithms nowadays is Direct Clustering Algorithm (DCA) and Rank Order Clustering (ROC). The comparison results are presented as shown below:

Products	Product Volume	Machine											
		M2	M10	M5	M4	M7	M12	M9	M1	M8	M6	M11	M3
A	2500	4	3	2					1				
D	1000	3	0	2					1				
E	800				4	0	3	2	1				
F	5000				5	3	4	2	1				
B	500								1	0	2	3	4
C	800								1	2	3	4	5

(A) EDM method

Products	Product Volume	Machine											
		M8	M3	M6	M11	M1	M7	M4	M9	M12	M5	M10	M2
C	800	2	5	3	4	1							
B	500	0	4	2	3	1							
F	5000					1	3	5	2	4			
E	800					1	0	4	2	3			
A	2500					1					2	3	4
D	1000					1					2	0	3

(B) DCA method

	Product Volume	Machine											
		M1	M2	M5	M10	M6	M11	M3	M8	M9	M12	M4	M7
A	2500	1	4	2	3								
D	1000	1	3	2	0								
C	800	1				3	4	5	2				
B	500	1				2	3	4	0				
F	5000	1								2	4	5	3
E	800	1								2	3	4	0

(C) *ROC method*

Figure 12. Comparison result betwen EDM, DCA and ROC method

Although the GT efficiency of the proposed method is not superior to the other two methods (75%), a special machines arrangement (MC1) mentioned in **Step 06** is considered outstanding and useful for a multi-routing approach that optimized material flow workflow. The layout of machinery was reasonable, ensuring the lead time as limiting the transportation time between cells together. Through reducing the cost of material handling can help reduce costs and resources for production to improve profits for businesses

5. Conclusion

From the results of the study, it can be concluded that the proposed Euclidean distance method is really effective in the layout design of multi-route plans to consider the volume of product - one of the important factors in the layout design field while some traditional methods have not been considered. By optimizing the performance of the layout, it minimizes movement during production, reducing transportation time and smoothing the product workflow. The proposed method has also been proven to be effective and can compete with other floor design methods (Manash Hazarika, 2016).

Reference

Aalaei, Amin; Davoudpour, Hamid (2017). A robust optimization model for cellular manufacturing system into supply chain management. International *Journal of Production Economics*. 183: 667–679

Dokmanic, Ivan; Parhizkar, Reza; Ranieri, Juri; Vetterli, Martin (2015). "Euclidean Distance Matrices: Essential theory, algorithms, and applications". *IEEE Signal Processing Magazine*. Page: 12–30.

Flanders, R.E (1925). Design, manufacture, and production control of a standard machine. *Transactions of ASME*, Vol. 26, 691-738.

Ivanov E K (1968). *Group Production Organization and Technology* (Book). Business Publications Ltd.

Inman, R. Anthony; Helm, Marilyn (2006). Encyclopedia of Management. Detroit*, MI: Gale Cengage Learning*. pp. 72–78

J. Saeed, M. Ameli and J. Arkat, (2008), Cell Formation with Alternative Process Routings and Machine Reliability Consideration, *The Int. J. Adv. Manuf. Technol.,* vol. 35, pp. 761–768.

Kumar, C. S., & Chandrasekharan, M. P. (1990). Grouping efficacy; a quantitative criterion for goodness of block diagonal forms of binary matrices in group technology. *International Journal of Production Research*, 28, 233–243.

Liker, Jeffery (2004). *The Toyota Way*. New York: McGraw Hill. pp. 31, 96–101.

Morgan, J.M. (2006). *The Toyota Product Development System*. New York: Productivity Press. p. 97.

ManashHazarika (2016). A Heuristic Approach for Machine Cell Formation Problems with Alternative Routings. *Procedia Computer Science* Volume 89, , Pages 228-242

N.Hyer, U Wemmerlöv (2002). Reorganizing the Factory: Competing through Cellular Manufacturing, First ed., *Productivity Press*, New York, NY.

Nikola Suzić et al (2012). Customizing Products through Application of Group Technology: A Case Study of Furniture Manufacturing. *Journal of Mechanical Engineering*

Amruthnath, Nagdev; Gupta, Tarun (2016). Modified Rank Order Clustering Algorithm Approach by Including Manufacturing Data. *IFAC-Papers OnLine*. 49 (5): 138–142.

Richtsmeier JT, and Lele S (1993). *A coordinate free approach to the analysis of growth patterns: models and theoretical considerations*. Biol. Rev. 68:381411.

Tamal Ghosh, Pranab K Dan (2011). An Effective Machine-Part Grouping Algorithm ton Construct Manufacturing. *Proceedings of Conference on Industrial Engineering (NCIE)*

Weyenberg, G., & Yoshida, R. (2015). Reconstructing the phylogeny: Computational methods In Algebraic and Discrete Mathematical methods for modern Biology (pp. 293-319). *Academic Press*.

Wafik Hachicha, Faouzi Masmoudi & Mohamed Haddar (2007). Cells Formation of machine groups and part families in cellular manufacturing systems using a correlation analysis approach. *Springer-Verlag London Limited 2007*

INTEGRATED FARM MODEL FOR SUSTAINABLE DEVELOPMENT IN BINH THUAN PROVINCE, VIET NAM

Trang Le Thi Kim[1], Wu C. Kenny[2], Phong Ho Thanh[3] and Wang Chia-Nan[4]

[1] *Le Thi Kim Trang, Business Intelligence School- National Kaohsiung University of Science and Technology, trangltk.hcm@gmail.com*

[2] *C. Kenny Wu*, *Business Intelligence School- National Kaohsiung University of Science and Technology, ckwu@nkust.edu.tw*

[3] *Ho Thanh Phong*, *Logistics and Supply chain Management Department- Hong Bang International University, phonght@gmail.com*

[4] *Chia-Nan Wang*, *Industrial Engineering and Management Department- National Kaohsiung University of Science and Technology,* cn.wang@nkust.edu.tw

Abstract. This research aims to develop a business model that combines traditional agriculture production of dragon fruit with rural tourism in Binh Thuan Province, Vietnam, by creating additional sources of revenue to increase farmers' income. A Linear Programming (LP) method is constructed with multi-objective designs of a weekly plan affected by seasonal factors such as peak season and off-season within a year. The result demonstrates a detailed schedule for farming and harvesting two types of dragon fruit. As a result, farmers have better manage production activities and define open time for tourist activities by a year. The outcome of this study showed that farmers' revenues was improved for more than seven hundred million VND in Binh Thuan due to increased tourist activities while planting dragon fruits. Therefore, farmers should establish a strong collaboration with the local community to build sustainable development of rural tourism and dragon fruit production in Binh Thuan province.

Keywords: Integrated farm management; tourism strategy; agritourism planning; production planning; agriculture sustainable development

1. Introduction

Vietnam is a country with a long tradition of agricultural production, especially growing and supplying a variety of agricultural products such as rice, fruits ... in the domestic and world markets. In particular, the export of dragon fruit is leading in 25 countries and territories with the largest area and production of dragon fruit in Asia. With favorable climate and soil characteristics, Binh Thuan province has been mentioned as the "capital" of dragon fruit which is leading area of dragon fruit production (30.000 ha, 2020). Each year, Binh Thuan supplies more than 600,000 tons of dragon fruit to the market (according to the statistics of the Ministry of Agriculture, 2020), which are grown at 10 districts in the area. Binh Thuan dragon fruits have branches develop strongly and take some special characteristics: good looking skin, strong meat and sweet. Despite of the advantage in natural conditions as well as the history of fruit development in the area, the fruits have faced some problems. The first problem is that the production cost of dragon fruit is still high which lead to reduce the competition in the domestic and international market. There is no production planning of fruits in the area. The additional costs will be occurred in the off-season related to the plant growing, electricity systems, irrigation systems and electricity cost. Farmer must spend from 25-30 million VND per 1,000 planting leads to produce more than 3.5 tons of dragon fruit. Moreover, the fruit have unstable the output of fruit and unvaried the fruit' consumption method because of depends mainly on Chinese markets. Almost, after harvesting, the dragon fruits in Binh Thuan are packed and consumed mainly in the form of fresh fruits in the market. The lack of the combination with tourism to develop its potential is another limitation of the current dragon fruit's farm in the area. Finally, the producer is now not well practice of fruit production by less understanding of ecosystem service as relationships between organisms and environment. For those reason, dragon fruit cannot promote the potential and expand in the region.

2. Literature review

(Paam, 2016) presents the review of planning models to optimize the Agri-fresh food supply chain for loss minimization. The studies focus on four main research steams related to modeling approaches, functional area, objective functions, solution approaches which help analyze and give a solution for minimization in the fruits and vegetables supply chain. As the modeling approach, there are two mains' models for agri-fresh food supply chain including deterministic and stochastic. Mathematical models have finding by four common types of decision variables as production, harvest, inventory and distribution. To continue, the objective function is categorized in two groups of single- objective and multi- objective combine with determined the type of objective functions in the research. Then, finding the methodologies such as heuristics algorithms, simulation tools, goal programming to solve the models is the next step. Moreover, the studies classified two other streams on agri-fresh food supply chain including: product type/country and case study.

(Villalobos, 2012) presented an integrated tactical planning model for production and distribution of fresh product. The decisions obtained are based on not only traditional factor such as price estimation, resource availability but also price dynamic, transportation, inventory cost.(Graves & Willems, 2005)indicates the design for optimizing the supply chain configuration for new products. The research aims to minimize the total supply chain cost involved goods sold cost, safety and pipe stock cost, from vendors to supply a certain raw materials, alternative machines or processes to manufacture the assembly, and alternative shipping modes to deliver goods to final customer.

Jose Vicente Filho (2002) used LP model to design a production planning for lily flower with objective to maximize the farm's total contribution margin.

(Liu & Yen, 2017) A conceptual framework for Agri-food tourism an Eco-Innovation Strategy in small farms. The purpose of this paper is to build the integrated model of small-scale farms as agricultural and tourism linkage in the Central Mountain Range of Taiwan. This case study mentions to optimize relationships which take improvement in both the upstream and downstream sides of this services. Due to sustainable development, the practices of farmer must consider several factors such as cultural, social and environmental systems.

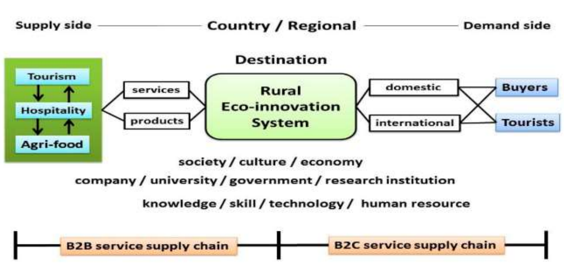


Figure 1. A proposed conceptual framework for agri-food tourism as eco-innovation strategy

(Bien & Thanh, 2015) illustrates the relationship between Tourism Development and Agricultural Commercial Production in Uong Bi City, Quang Ninh.

This result of the study is shows by making analysis in agricultural production related to product variety, productivity and market consumption combine with tourism activities such as tourism demand, tourism products, diversification, culture and local policy.

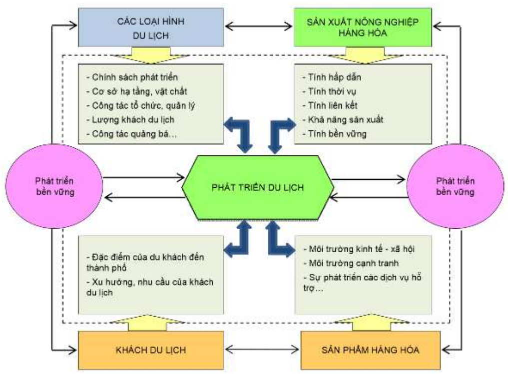


Figure 2. Relationship between tourism development and agri-food production

3. Methodology

3.1. Agri-tourism analysis

Agritourism is a new type of tourism which creates products based on resources of agriculture activities. It is designed to combine agriculture and tourism activities; attract tourists for sightseeing, farming and harvesting; bring high income for farmers; discover new entertaining and physical training means; help visitors to get closer to the nature and experience rural lifestyle. By matching natural factors with cultural values in rural area, Argi-tourism not only preserves the country values but also maintains traditional jobs in Binh Thuan province.

There are records for developed countries like Japan, Taiwan applying agri-tourism projects and earning profits from them. Vietnam, a tropical country with good weather for farming activities, may benefit from both tourism and agriculture industries. Therefore, Agritourism should be studied as a new approach to utilize Binh Thuan resources. Current agricultural production with various tourism types provides good supply for agritourism. There are many existing tourism products in Vietnam with agricultural and cultural values across the country attracting local and foreign visitors such as: Duong Lam commune tours (Hanoi), Moc Chau farm tours (Son La province), Tra Que vegetable village tour (Quang Nam province), Da Lat farm tour (Lam Dong province), Can Gio ecological tour (HCM), Mekongdelta eco-tours.

In this case, Binh Thuan dragon fruit need to have great a strategy to develop their potential. The combination between agricultural production and tourist services as an example that help improve current's their business as well as regional market-oriented agricultural economy. The performance of this integrated model is focused on improving the production activities and expanding farm's services. The destination has also changed significantly by providing more services related to visit the planting area by learning about knowledge of dragon fruit production as well as local activities and cultures, sightseeing, take a picture and so on. The direct sales in farm are a new form consumption in market. After visiting the farm, tourists can immediately choose and harvesting the product by themselves. It is interesting activities and make a good experience for them. By attracting customer in long- term, the farm should certainly mention some factors such as natural landscape, comfortable feelings.

3.2. Mathematical model

3.2.1. Assumption

The production planning model with LP – multi objectives for planting dragon fruit companies is chosen. This mode is designed based on weekly records to provide company leaders with farming statistics. Random factors like productivity and weather condition are assumed to be determinants**.**

3.2.2. Set description:

No	Set	Index	Set description	Length
1	T	$t \in T$	time period	48 weeks
2	I	$i \in I$	type of product	2 products
3	J	$j \in J$	Set of area	29 fields

3.2.3. Input parameters:

Parameters	Descriptions
CAP_j	Number of planting slot in the field j
$PRICE_{it}$	Price of product type i at time t
$CGRW_{it}$	Cost for planting product type i at time t on one planting slot
$HOLD_i$	Cost for storing product type i
$VIST_j$	Visitor amount in the field j
MinInv	Minimum inventory 0
MaxInv	Maximum inventory at 20 ton
$Demand_{it}$	Demand of product type i at time t

3.2.4. Decision variables:

No	Decision Variables	Descriptions
1	X_{ij}	The product type i in the field j
2	HAR_{ijt}	Number of product type i in the field j at time t
3	GRW_{ijt}	Number of product type i from growing flower in the field j at time t
4	$AVAI_{ijt}$	Number of remain plant in the field j at time t
5	QAN_{it}	Number of harvesting product type i at time t
6	$PROD_i$	The productivity of product type i
7	IW_{it}	The amount of inventory of product type i at time t
8	$CLOSE_{jt}$	= 1, if slot j is not open for visiting at time t
9	$SELL_{it}$	Number of product type i sale at time t

3.2.5. Objective function

The first objective is minimized the total planting cost as minimizing electricity consumption cost:

$$\sum_{i}\sum_{i}\sum_{t} CGRW_{it} * GRW_{ijt}$$

The second objective is minimized number of visitors who cannot enter the farm variable:

$$\sum_{j}\sum_{t} CLOSE_{jt} * VIST_{j}$$

The last objective of this model is minimized inventory cost:

$$\sum_{i}\sum_{t} HOLD_{i} * IW_{it}$$

3.2.6. Constraints

Constraint 1: Number of harvesting slots = Number of planting slots before 7 weeks

$$HAR_{ijt} = GRW_{ijt-7} \quad \forall i,j,t$$

Constraint 2: Number of slot available for each farm

$$AVAI_{jt} = CAP_j - \sum_{i}\sum_{k\le t} GRW_{ijk} + \sum_{i} HAR_{ijt} \quad \forall j,t$$

Constraint 3: Planting quantity

$$GRW_{ijt} \le AVAI_{jt} \quad \forall i,j,t$$

Constraint 4: Identify planting farm type

$$GRW_{ijt} \le X_{ij} * BigM$$

Constraint 5: Each farm produces one type of product only

$$\sum_{i} X_{ij} = 1$$

Constraint 6: Harvesting quantity

$$QAN_{it} = \sum_{j} HAR_{ijt} * PROD_i \quad \forall i,t$$

Constraint 7: Inventory

$$IW_{it} = IW_{it-1} + QAN_{it} - SELL_{it} \quad \forall i,t$$

Constraint 8: Inventory Limitation

$$Min\ inventory \le IW_{it} \le Max\ inventory$$

Constraint 9: Demand

$$SELL_{it} = demand_{it} \quad \forall i,t$$

Constraint 10: Number of harvesting plant for opening tourism activities status

$$\left(1 - CLOSE_{jt}\right) \le \sum_{i} HAR_{ijt}$$

4. Model result

After using CPLEX software to implement the model, the optimal solution for the integrated farm based on three objectives is showed. The result demonstrates a plan for planting 2 types of dragon fruit – white and red in 29 farms. Based on the proposal, farmers can control, manage their farm and their products, and identify which product produces in which farm. The graph below indicates the level of planting white and red dragon fruit peak season and off-season.

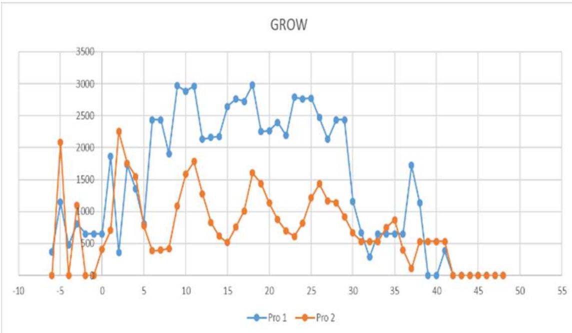


Figure 3. The area of planning chart

The chart below shows the profit reached more than seven hundred million VND. The result describes the decrease of production cost and the raise of the profit for farm's activities due to increased tourist activities while planting dragon fruits.

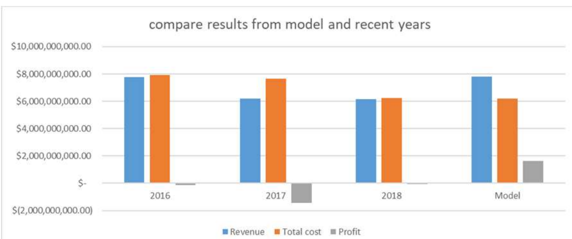


Figure 4. The economic analysis chart

The production planning also describes the time for tourist activities, especially as the tourist opening time as peak season coincides with summer season from week 17 to week 36 (May – September). This time is tourist peak season which receives large number of visitors to bring the profits by direct sales at farm.

Figure 5. The number of opening times of planting area chart

5. Conclusion

The proposed model of the study is used for the producer who want to develop the Dragon fruit production in this area. This research aims to improve the current dragon fruit production in Binh Thuan Province. LP method is applied in this model to solve multiple objectives. The achieved results come up with a weekly detailed production plan for peak season and off-season durations, which is used to manage tourists' activities. By applying this model result into production, farmers can not only save planting costs, earn high income, utilize resources but also create a new way of business as agri- tourism. However, there are some limitations observed in this model, including random factors such as demand, product prices, weather conditions. Then, with the current resources of 50.000 ha in area, the farm can adapt to increase of demand in future.

Producers should draw more studies on market demand and collect information about customer habits to attract more visitors related to network with tourist agencies, product quality standards, facilities investment and increasing brand awareness. For Binh Thuan local community should implement policies to support dragon fruit production involved encourage more investments to dragon fruit production firms, invite more actors in the chain as co-operatives, companies and organizations in agri-tourism industry and create a sustainable growth by matching product quality standards with environmental factors.

Agricultural tourism is a new type of business with great potential for development in the future. The combination of agricultural production and tourism as promoting products brings many benefits to farmers as well as economic growth and reputation in the local. This study is to pilot on a small farm in the area. Hopefully, this model will be applied successfully and expanded on a large scale in Binh Thuan province.

Reference

Paam, P., Berretta, R., Heydar, M., Middleton, R. H., García-Flores, R., & Juliano, P. (2016). Planning models to optimize the agri-fresh food supply chain for loss minimization: a review. *Reference Module in Food Science*, 19-54.

Ahumada, O., Villalobos, J. R., & Mason, A. N. (2012). Tactical planning of the production and distribution of fresh agricultural products under uncertainty. *Agricultural Systems*, 112, 17-26.

Graves, S. C., & Willems, S. P. (2005). Optimizing the supply chain configuration for new products. *Management science*, 51(8), 1165-1180.

Caixeta-Filho, J. V., van Swaay-Neto, J. M., & Wagemaker, A. D. P. (2002). Optimization of the production planning and trade of lily flowers at Jan de Wit Company. *Interfaces*, 32(1), 35-46.

Liu, S. Y., Yen, C. Y., Tsai, K. N., & Lo, W. S. (2017). A conceptual framework for agri-food tourism as an eco-innovation strategy in small farms. *Sustainability*, 9(10), 1683.

Biên, T. X., Thành, N. X., & Hải, Đ. N. (2015). NGHIÊN CỨU MỐI QUAN HỆ GIỮA PHÁT TRIỂN DU LỊCH VỚI SẢN XUẤT NÔNG NGHIỆP HÀNG HÓA TRÊN ĐỊA BÀN THÀNH PHỐ UÔNG BÍ, TỈNH QUẢNG NINH. J. *Sci, 13(7)*, 1222-1231.

From Agriculture to Agr-tourism Report Central, Okanagan Economic Development Commission, (2012).

Agriculture and tourism linkages in Pacific Island Countries on a scoping mission in Samoa and Tonga Report- FAO organization (2012)

Agri-tourism and the Farmer as Rural Entrepreneur, Wilfred Laurier University, Waterloo, Ontario in April (2010)

Agri-tourism: As a new element of country planning Report (2007)

Official website of Ham Thuan Nam district- Vietnam- http:// hamthuannam.binhthuan.gov.vn/

Official website of Department of Statictics of Binh Thuan Province- Vietnam- http:// cucthongke.com.vn/

APPLICATION OF MACHINE LEARNING AND OPTIMIZATION TECHNIQUE ON DEMAND AND SUPPLY PLANNING

Le Dinh Quoc Huy[1] and Phan Nguyen Ky Phuc[2]

[1] *International University of Ho Chi Minh City, Vietnam, ldqhuy.iseyu@gmail.com*

[2] *International University of Ho Chi Minh City, Vietnam, pnkphuc@hcmiu.edu.vn*

Abstract. Supply Chain Planning including Demand Planning and Supply Planning is one of the most critical steps which decides the overall performance of supply chain. Especially, new century of big data extracted from complex behaviors of customers and intense competition between players in the same fields has motivate many companies make investment to improve supply chain planning process. At the same time, the potential growth of Machine Learning algorithms and optimization techniques has allowed experts and researchers to apply these techniques in sales forecasting as well as make approriate inventory plan to satisfy the market demand. This paper aims to fullfill three main goal. Firstly, this paper proposes a sales prediction model using Gradient Boosting method, paying attention to products' perishability and profitability in feature engineering process. Secondly, an optimal inventory ordering policy shall be designed with the consideration in discount quantity policies and remaining shelf-life requirement from customers. Finally, some solutions from the perspectives of business acumen will be proposed in addition to engineering solutions so as to provide a full picture of supply chain risk management for perishable goods.

Keywords: demand planning; gradient boosting; optimal inventory policy; quantity discount; perishable goods.

1. Introduction

Demand planning is the process of applying methodologies and information technologies, utilizing input data from relevant departments to trigger demand forecasts for the future. The aim is to accelerate the flow of raw materials, materials and services beginning with the suppliers through transforming to products in the company and to their distribution to their final consumers (Vickova, Patak 2015). The error from demand forecasts of demand planning process are directly related to required safety stocks, while frequent adjustments of demand forecasts can lead to dramatic changes in plans (Stadtler 2005).

Thus, demand planning is one of the critical stages in supply chain management. The outcome of this function has the large impact on inventory situation, purchasing decision, sales performance as well as cycle service level. Thanks to this step, the future sales pattern is explored which can be considered as the solid foundation for an effective supply chain.

To fullfil the requirement of customers, supply planning is the next step of demand planning. This stage includes decision-making processes relating to purchasing decision and inventory level management. One of the biggest challange for supply planners is how to achieve high service level while minimize types of cost for the business. As a result, excessive stock and out-of-stock are two main serious problems for business in daily operation

Out-of-stock can lead to lost sales which are considered as risk not only for today, but also for future. In such competitive market, customers can shift to many other distributors or purchase sustitute item when the current choices do not satisfy their expectation. In some case, out-of-stock will make the business violate their commitment which incurrs compensation and damage company's reputation.

On the other hand, excessive stock reveals the potential problem for cost management such as incurred cost for storage, handling, labor, insurance... Especially, for product which has limited shelf life and special storage requirement, there exists high risk of products' quality reduction and degradation. In addition, holding large amount of inventory for a long time could reduce the flexibility of working capital that can be used for more important and urgent purposes.

Demand Planning and Supply Planning have never been easy tasks. Nowadays, they become even much more difficult with the changes of Supply Chain's nature. Supply chain are more and more complicated with the participation of new players that create the fierce competition between competitors. In addition, the increasing demand of the market leads to the complication of supply chain network and higher dimension of supply chain data. As a result, the key to be the winner in this competition is how to utilize the huge amount of information and successfully deal with uncertainty in such fluctuating demand of market.

The main objective of this research is to optimize the demand and supply planning process. After the forecasting sales data is calculated from the historical data using machine learning model. A decision-making process in order policy of review period will be conducted to ensure the minimal inventory cost while satisfy the customer demand at target service level. Finally, some suggestions from management perspective shall be given so as to improve the demand and supply planning process at the company.

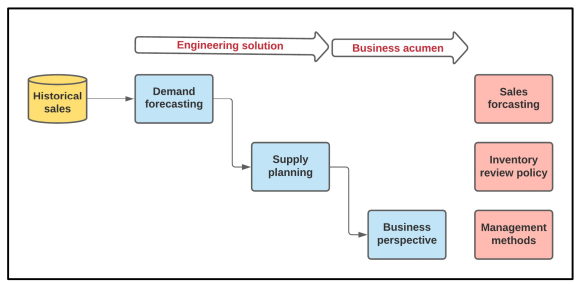


Figure 1: Objectives of the research

2. Engineering solution

2.1. Demand Planning

There are 2 datasets that are given as the input for the demand planning process:

- Product master data: Dataset show all basic
- Sales data: Dataset records the demand of each product in carton on a daily basis

The information and relationship between two datasets can be shown by entity-relationship diagram as below

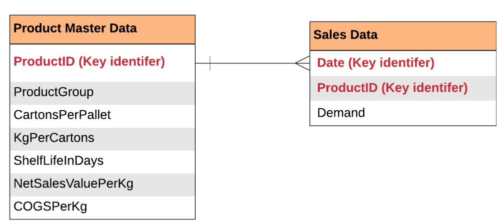


Figure 2: Entity-relationship diagram of demand planning data

2.1.1. Framework

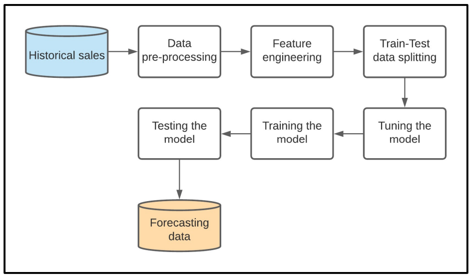


Figure 3: Demand Planning Framework

Data-preprocessing

In this step, some basic data processing steps will be conducted on the raw data as the input preparation for the forecasting model. These steps include:

- **Data consolidation:** Data from many source such as master data, transaction data, … will be joined together using key idenfiers of each data source
- **Data filtering**: Based on the available data of each object, only objects which have as much data as requirement could be selected to join the demand planning process. In other words, new products will not be considered in the scope of this research.
- **Outliers' adjustment:** Outliers accompanied with adnormal will be considered as "noise" and shall be adjusted to improve the overall performance of forecasting model.
- **Variable formatting:** Each variable should be labeled with approriate data class such as continuous variable/ categorized variable so that the machine can understand and learning meaningful features from input data.
- **Variation normalization:** With variables with large range or strong variation between objects, variation normalization could be considered as the solution to speed up the training process.

Feature engineering

Feature engineering is the process of using domain knowledge to extract features from raw data via data mining techniques. These features can be used to improve the performance of machine learning algorithms. Feature engineering can be considered as applied machine learning itself.

Especially for the demand planning case, some specific knowledge about the supply chain and product could be useful for creating meaningful variables so as to improve learning process of the machine which leads to the better forecasting performance.

Train-test data splitting

Full input data will be splitted with suitable ratio for the model training and testing process. There is some preferrable ratio that could be considered such as 60/40, 70/30, …

Model tuning

A machine learning algorithm always contains a set of parameters. Decision on the value of these parameters play as the critical part to the outcome of the model. For this problem, model tuning is the solution which use the computer programming to find the optimal set of parameters within a predefined range using the cross-validation technique.

Training and testing

After getting the optimal set of parameters, training and testing process will be conducted to find the solution and evaluate the performance of the model

2.1.2. Method considered:

Gradient boosting is a machine learning technique for regression and classification problems that produce a prediction model in the form of an ensemble of weak prediction models. This technique builds a model in a stage-wise fashion and generalizes the model by allowing optimization of an arbitrary differentiable loss function. Gradient boosting basically combines weak learners into a single strong learner in an iterative fashion. As each weak learner is added, a new model is fitted to provide a more accurate estimate of the response variable. The new weak learners are maximally correlated with the negative gradient of the loss function, associated with the whole ensemble. The idea of gradient boosting is that we can combine a group of relatively weak prediction models to build a stronger prediction model.

Input:

Data: $\{(x_i, y_i)\}_{i=1}^{n}$

Differentiable loss function: $L(y_i, F(x))$

Learning rate: $\nu \in [0;1]$

Algorithm:

Phase 1: Initialize model with constant value:

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$$

Phase 2: for m=1 to M

- Compute new Pseudo Residual (negative gradient)

$$r_{im} = -\left[\frac{\partial L(y_i, F(x))}{\partial F(x)}\right]$$

- Fit a regression tree to the r_{im} values giving terminal region R_{jm} for j=1… J_m
- For j=1… J_m compute

$$\gamma = argmin_\gamma \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma)$$

- Update new value

$$F_m(x_i) = F_{m-1}(x_i) + \nu \sum_{j=1}^{J_m} \gamma_i I(x_i \in R_{ij})$$

Some key hyperparameters in Gradient Boosting:

- Number of trees: Number of iterations
- Maximum nodes per tree: Number of splits it has to perform on a tree (starting from a single node).
- Shrinkage: Considered as learning rate, used for reducing, or shrinking, the impact of each additional fitted base-learner (tree). It reduces the size of incremental steps and thus penalizes the importance of each consecutive iteration.
- The minimum number of observations in trees' terminal nodes: Can be considered as a stop condition of algorithms where the machine will split each node until requirement about minimum number of observations in trees' terminal nodes is satisfied

2.1.3. Execution:

Data consolidation

From the given set of data, the forecasting data will be created by merging "Product Master Data" and "Sales Data" using key identifier "Product ID". The full dataset includes all information from two original datasets with variety of input variables for forecasting model.

Data filtering

The purpose of data filtering step is to remove new launch product from the dataset to ensure the accuracy and confidence of forecasting results.

To provide the confident and objective results, a train-test ratio of 67%-33% is chosen to split the data. In other words, 67% of available historical data will be chosen to partipate in training the forecasting model which will provide the sales forecast during the planning horizon. The sales forecast results will be compared with the test data to measure the forecast accuracy.

Within the scope of this research, the author choose the 30 months which approximate 300 days as the demand planning horizon. As a result, to meet all above requirements, each product should have the available sales data of at least 900 days. The others which do not have enough sales data will be removed from this model to participate later or join another model with shorter planning horizon

Outliers' adjustment

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. For fast-moving consumer goods, most of outliers in sales data can be explained by the sales program from the sales department or special events (such as disaster, pandemic,.. etc.) rather than the normal demand of the market.

Within the scope of the provided data, the forecasting model is built to provide the baseline sales forecasting value. For further research of this topic, this value can be adjusted by adding more objective factors such as sales incentives, sales promotion which will boost the demand or undesirable factors which will decrease the demand.

In descriptive statistics, box and whisker plot (also known as box plot) is one of popular type of chart used in explanatory data analysis. Box plots visually show the distribution of numerical data and skewness through displaying the data quartiles (or percentiles) and averages. Based on the visualization, the outliers of the dataset will be detected by determining of the observations which fall out of the outer fences of te boxplot chart.

After being detected, the outliers shall be replaced by Q2 – median value of observing attribute which represents for the demand market at normal condition. There is a special note that this outlier's adjustment process should be consider within sales data of each products instead of sales data of all products because each product has different demand characteristics which could lead to the different boxplot value and outliers determination.

Variation normalization

With big size of given data, saving time by speed up training process is necessary. There is two main methods that can be considered to make this training process faster including

- Scaling the input attribute using normalization method
- Increasing the learning rate

As discussed in previous part, gradient boosting is considered a gradient descent algorithm. An important parameter in gradient descent is the size of the steps which is controlled by the learning rate. If the learning rate is too small, then the algorithm will take many iterations (steps) to find the minimum which means that much time is spent in the training process. On the other hand, if the learning rate is too high, we might jump across the minimum and end up further away than when we started.

Thus, choosing small learning rate while scaling down the input value shall be the safe way to speed up the training time while not affect to the optimization level of the final result.

For this research, natural logarithm scaling, which is a popular scaling method, is used to normalize the "Demand" variables.

Feature engineering

Feature engineering is the process of using domain knowledge to extract features from raw data via data mining techniques. These features can be used to improve the performance of machine learning algorithms. Feature engineering can be considered as applied machine learning itself.

In other word, the more meaningful variables from feature engineering process are added into model, the more effectively the learning model could perform. Based on the features relating to shelf-life and stock movement from the historical data, the perishability risk label is defined for each product to categorize products with similar feature into groups using the rule: *"The slower stock movement cooperating with the shorter shelf-life could creating higher risk of perishability of product"*.

In addition, another categorizing process for products can be conducted using ABC classification based on profit which is an application of 80-20 rule.

Using these classification methods, from the perspective of engineering, the forecasting model shall be improved because more insightful variables are created and inputed. In addition, an approriate reconcilation plan with operation department and sales department shall be triggered to improve inventory health and sales performance from the perspective of management.

Perishability Penalty convention

Table 1: Shelf-life label

Shelf-life (in days)		Shelf-life Label
Lower fence	Upper fence	
0	90	Short
180	240	Medium
364	456	Long

Table 2: Stock movement Label

Cumulative contribution to daily sales		Stock Movement Label
Lower fence	Upper fence	
0	0.7	Fast
0.7	0.9	Normal
0.9	1	Slow

Table 3: Perishable Penalty convention

Shelf-life	Long	C1	C2	C3
	Medium	B1	B2	B3
	Short	A1	A2	A3
Perishability Risk		Slow	Normal	Fast
		Stock Movement		

Profit class convention

Table 4: Profit class convention

Cumulative contribution to daily profit		Profitability Label
Lower fence	Upper fence	
0	0.7	A
0.7	0.9	B
0.9	1	C

Model tuning

To execute the hyperparameters selection process, Grid Search, also known as parameter sweeping, is applied

Grid Search is one of the most basic and traditional methods of hyperparametric optimization. This method involves manually defining a subset of the hyperparametric space and exhausting all combinations of the specified hyperparameter subsets. Each combination's performance is then evaluated, typically using cross-validation, and the best performing hyperparametric combination is chosen.

Firstly, a set of possible value for each model tuning should be defined for Grid Search process as below:

Table 5: Sets of possible values for model tuning

Hyperparameter	Possible values
Number of trees	150, 200, 250, 300, 350, 400, 450, 500, 550
Maximum nodes per tree	1, 3, 5, 7, 9, 11
Shrinkage (learning rate)	0.1
The minimum number of observations in trees' terminal nodes	1, 5, 9, 11

As the tuning rule provided by Grid Search methods, a large number of training cases will be performed based on the number of elements inputed to set of possible values for each hyperparameter.

Thus, for this tuning process, there will 216 training cases that is created and evaluated. The evaluation metrics are used for this tuning process are root mean squared error (RMSE) and Coefficient of determination R-squared (R^2)

Root mean squared error determine the difference between the prediction value and the real value using following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

where:

y_i: The actual value of observation.

$\hat{y}_i$: The prediction value of observation.

R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Whereas correlation explains the strength of the relationship between an independent and dependent variable, R-squared explains to what extent the variance of one variable explains the variance of the second variable.

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_i(y_i - \hat{y}_i)^2}{\sum_i(y_i - \bar{y}_i)^2}$$

where:

y_i: The actual value of observation.

$\hat{y}_i$: The prediction value of observation.

RSS: Sum of squares of residuals

TSS: Total sum of squares

The following plotted charts show the result of tuning process, using two explained performance metrics.

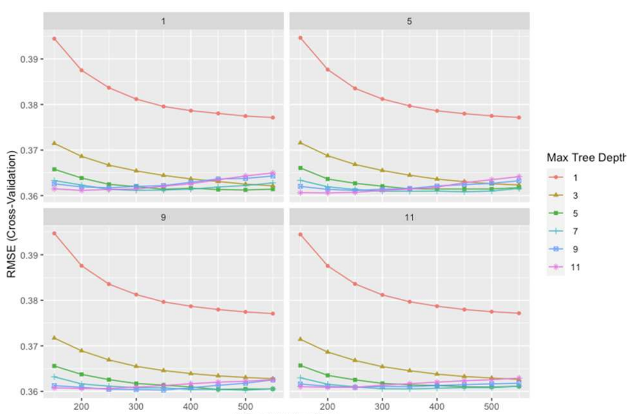


Figure 4: RMSE in model tuning process

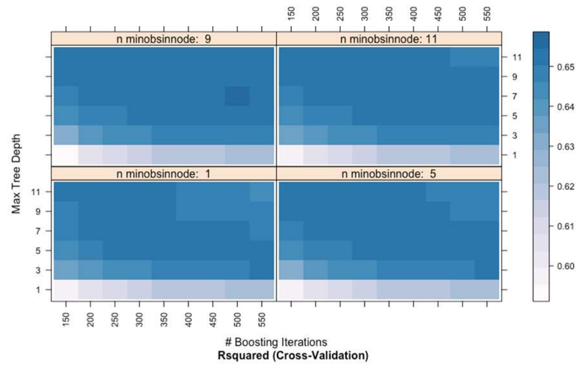


Figure 5: R-squared in model tuning process

In general, it does not have much difference in the result provided by four inputed values of minimum number of observations in trees' terminal nodes

From the RMSE chart, it can be seen that acrossing four inputed values of minimum number of observations in trees' terminal nodes, the common trend is that the RSME initially keep decreasing with respect to the increase of number of trees until it reaches the optimal point. Then, the RMSE starts to grow again despite the rise of number of trees.

Comparing between inputed value for maximum nodes per tree, there exists the significant improvement when increasing this value from one to three. The enhancement gap gradually becomes smaller and less correlated to the rise of maximum nodes per tree. Concurrently, when increasing maximum nodes per tree, the required number of trees that need to be created to reach the optimal point reduces.

From the R-squared chart, the model can provide the coefficient of determination up to 66% - represented the darkest region, which means that 66% of variance in the dependent variable that can be explained by the independent variable.

In addition, this tuning process also show the importance of each feature to the forecasting model using normalized R-squared metrics. PerishabilityRisk variables, which are created from feature engineering process, have outstanding performance in comparision with other attributes. This reveals the success of feature engineering process and contribution of this step to the overall performance of the forecasting model.

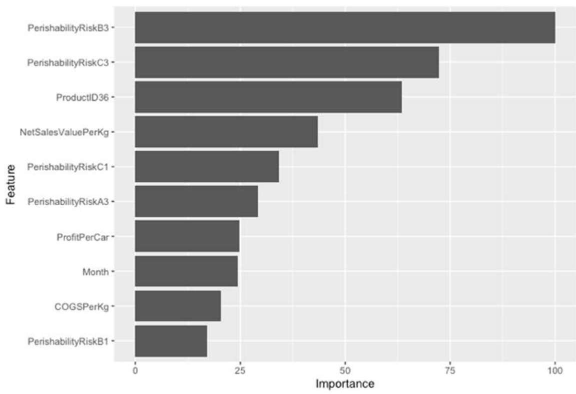


Figure 6: Top 10 of the most important features

Training and testing

The testing data, along with the optimal set of hyperparameters from the model tuning process, is inputed into the forecasting model in order to get the prediction value.

The performance metrics used to evaluate the forecasting result are Mean Absolute Percentage Error (MAPE) and Forecast Bias (FB) – forecasting performance metrics that are popular used for supply chain management in the real world.

MAPE is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual or observed value minus the forecasted value.

$$\mathrm{MAPE} = \left|\frac{(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})}{\mathrm{y_i}}\right|$$

where:

y_i: The actual value of observation.

$\hat{y}_i$: The prediction value of observation.

FB is a tendency for a forecast to be consistently higher or lower than the actual value.

$$\mathrm{FB} = 1 - \left|\frac{\mathrm{y_i}}{\hat{\mathrm{y}}_\mathrm{i}}\right|$$

where:

y_i: The actual value of observation.

$\hat{y}_i$: The prediction value of observation.

2.2. Supply Planning

As discussed in previous part, supply planning is the successor process of demand planning. This process uses the output of predecessor process as the input to define the optimal inventory model which meet the requirement about cost minization and order fill rate, considering the order quantity and perishability constraint. Products are divided into two groups with respect to two discount methods: All Unit Discount and Incremental Discount. In addition, for this problem, demand of each item is assumed to follow Normal Distribution where mean and standard deviation are parameters describing the characteristics of the demand.

2.2.1. Framework

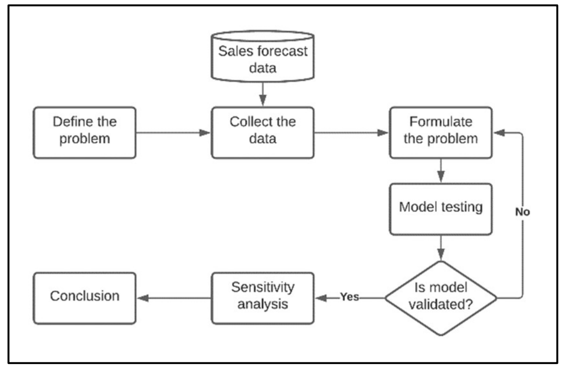


Figure 7: Supply Planning Framework

Define the problem

From the problem statement, objectives and constraints for the supply planning model should be described. Based on the inventory review policy and the way of calculation, decision variables and parameters shall be defined for next steps

Collect the data

From the requirement about types of paramaters that need to be supplied as the input of the models, the data collection is the next step. With the scope of this research, main data source used in the supply planning process is the sales forecast data which is generated from previous process.

Formulate the problem

Based on objectives and types of constraint from "Define the problem", a mathematical model shall be built which represents our problems by the language of operation research. In addition, the data class of each type of variables and parameters should be defined carefully so that feasible and applicable results can be found from the model.

Model testing

Mathematical model will be converted to the programming language depending to the tool that is chosen to test the model. Through this test, we can got the final solution which can be used to be validated with given constraints of the problem. Passing this validation steps, the solution shall be considered as the optimal solution of the problem. Otherwise, problem formulation step should be revisited to ensure the mathematical model is adjusted approriately.

Sensitivity analysis

For more insightful information, which is the foundation for further suggestion, a deep dive analysis called sensitivity analysis is necessary.

Through this process, the small adjustment of parameters will be made and impact of these changes on result of decision variables and objective function will be observed to extract informative insight about the model.

2.2.2. Method considered

Types of quantity discount strategy

There are two main quantity discount strategies that are applied widely by suppliers. All-unit discount where a common discounted price is applied for all units in the order where the size of order satisfy the lower fence of discount quantity range. Incremental discount where the discount price is applied only for the number of units that are purchase above the threshold.

While having some differences in purchasing cost calculation, both these two methods persuade the customer to purchase as much as possible so as to decrease cost paid for per unit purchased.

Linear programming

Linear programming may be defined as the problem of maximizing or minimizing a linear function which is subjected to linear constraints. The constraints may be equalities or inequalities. The optimisation problems involve the calculation of profit and loss. Linear programming problems are an important class of optimisation problems, that helps to find the feasible region and optimise the solution in order to have the highest or lowest value of the function.

Approximation of non-linear function

From the fact that all the input for the linear programming including objective functions and constraints should be "linear", all non-linear relationships should be approximate to linear ones to ensure the requirement of linear programming.

The basic knowledge about plane geometry shows that on a plane, there is one and only one line segment which can connect two points whose coordinators on the plane are given. Applying that foundation, the idea for approximation of non-linear function is to subset the non-linear graph into many line segments based on the coordinators of pairs of points which are calculated from the non-linear function.

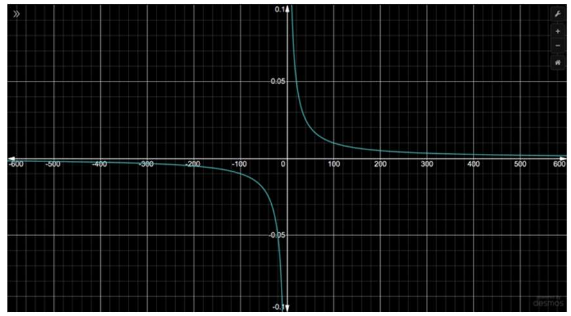


Figure 8: Graph of non-linear function

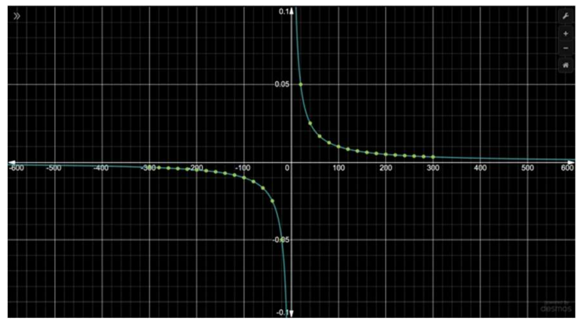


Figure 9: Set of points belongs to non-linear graph

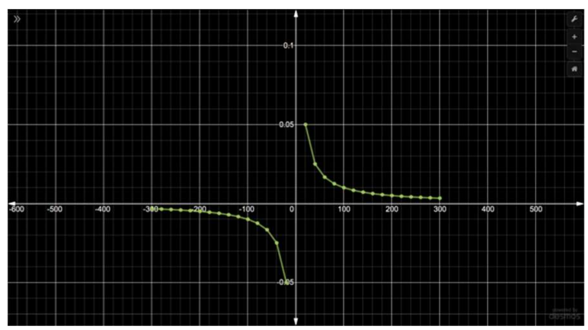


Figure 10: Approximation of non-linear function

The linear approximation follows process below

Phase 1: Determine line segment equation from pairs of points which are determined by the non-linear function

Step 1.1: Calculate coordiantors for a set of pairs of points from the non-linear function.

Step 1.2: From calculated coordinators, define the equation of line segments passing each pair of points: In the xy plane, the linear equation of line segment that passes two points A and B have the form:

$$\frac{x - x_A}{x_B - x_A} = \frac{y - y_A}{y_B - y_A}$$

In slope-intercept form:

$$y = \frac{y_B - y_A}{x_A - x_B}x - \frac{x_A(y_B - y_A)}{x_B - x_A} + y_A$$

From step 1.2, line segment equation for each interval has been determined

Phase 2: Approximate the value $f(x_i)$ of non-linear function $f(x)$, given input x_i

Step 2.1: Determine the interval (from phase 1) that the given point x_i belongs to

Step 2.2: Plug into the approriate line segment equation to get the approximated value of $f(x_i)$

2.2.3. Execution

Mathematical model

Sets:

- **I** Set of products
- $\mathbf{I_1}$ Set of products using All-unit discount
- $\mathbf{I_2}$ Set of products using incremental discount
- **J** Set of discount level
- **X** Set of linear approximation process
- **Y** Set of input for linear approximation process

Indices:

- **i** Product index
- **j** Level of discount index
- **x** Linear approximation process index
- **y** Linear approximation input

Parameters:

- $\mathbf{a_{ij}}$ Purchasing price for item i using discount level j
- $\mathbf{z_i}$ Service index of product i
- $\mathbf{g_i}$ Value of standard normal loss function
- $\mathbf{k_i}$ Profit of product i
- $\mathbf{u_{ij}}$ Upper fence of discount level j applied for product i
- $\mathbf{l_{ij}}$ Lower fence of discount level j applied for product i
- $\mathbf{m_i}$ Mean daily demand of product i
- $\mathbf{s_i}$ Standard deviation daily demand of product i
- **R** Interest rate used for calculating holding cost
- **H** Planning Horizon
- $\mathbf{o_i}$ Fixed ordering cost applied for product i
- $\mathbf{n_i}$ Shelf-life of product i
- $\mathbf{b_{ixy}}$ y^{th} input from item i for linear approximation process x^{th}
- $\mathbf{c_{ixy}}$ Output from y^{th} input of item i for linear approximation process x^{th}
- $\boldsymbol{\beta_{ixy}}$ Intercept of linear equation extracted from input y^{th} and $(y+1)^{th}$ of linear approximation process x^{th}
- $\boldsymbol{\alpha_{ixy}}$ Slope of linear equation extracted from input y^{th} and $(y+1)^{th}$ of linear approximation process x^{th}
- **BigM** Big number

Decision variables:

- $\mathbf{T_i}$ Review period of product i
- $\mathbf{Q_i}$ Total order quantity of item i during the horizon
- $\mathbf{q_{ij}}$ Order quantity of item i at discount level j
- $\mathbf{E_{ij}}$ Place order for item I at discount level j
  - $E_{ij} = 1$: Place order for item I at discount level j
  - $E_{ij} = 0$: Do not place order for item I at discount level j
- $\mathbf{p_{ij}}$ Purchasing cost paid for item i at level j
- $\mathbf{P_i}$ Total Purchasing cost paid for item i during the horizon
- $\mathbf{Z_{ixy}}$ Binary decision variable to determine the applicable range that input belongs to.
  - Z_{ixy} =1: Provided inputed input for item i belongs to the ranges $[b_{xy}, b_{x(y+1)}]$ for all linear approximation process x
  - Z_{ixy} =0: Provided inputed input for item i does not belong to the ranges $[b_{xy}, b_{x(y+1)}]$ for all l linear approximation process x
- $\mathbf{t_{ixy}}$ Temporary variable recording T_i for the input range $[b_{xy}, b_{x(y+1)}]$ for linear approximation process y
- $\mathbf{r_{ixy}}$ Semi-temporary variable recording result for the input range $[b_{xy}, b_{x(y+1)}]$ for linear approximation process y for T_i
- $\mathbf{f_{ixy}}$ Temporary variable recording result for the input range $[b_{xy}, b_{x(y+1)}]$ for linear approximation process y for T_i
- $\mathbf{F_{ix}}$ Result of T_i from linear approximation process y

Objective function:

Original objective function
Minimize

$$\sum_i P_i \frac{H}{T_i} + \sum_i \frac{H}{T_i} o_i + \sum_i \left(\frac{Q_i}{2} + z_i s_i \sqrt{T_i}\right) \frac{P_i}{D_i T_i} * R * H + \sum_i k_i \frac{H}{T_i} g_i s_i \sqrt{T_i}$$

Linearized objective function

$\sum_i P_i * H * F_{i1} + \sum_i H * F_{i1} * o_i + \sum_i \left(\frac{Q_i}{2} + z_i * s_i * F_{i2}\right) \frac{P_i}{D_i} * F_{i1} R * H + \sum_i k_i * H * F_{i1} * g_i * s_i * F_{i2}$

Constraint:

Market demand constraints

$1 \le T_i \le 0.3 n_i \quad \forall i \in I$ (1)

$\sum_j q_{ij} = Q_i \quad \forall i \in I$ (2)

$\sum_j p_{ij} = P_i \quad \forall i \in I$ (3)

$Q_i \ge m_i T_i \quad \forall i \in I$ (4)

Discount order quantity constraints

$Q_{ij} \le u_{ij} + BigM * (1 - E_{ij}) \quad \forall i \in I, j \in J$ (5)

$Q_{ij} \ge l_{ij} E_{ij} - BigM * (1 - E_{ij}) \quad \forall i \in I, j \in J$ (6)

$Q_{ij} \le BigM * E_{ij} \quad \forall i \in I, j \in J$ (7)

$p_{ij} \le q_{ij} * a_{ij} + BigM * (1 - E_{ij}) \quad \forall i \in I_1, j \in J$ (8)

$p_{ij} \ge q_{ij} * a_{ij} - BigM * (1 - E_{ij}) \quad \forall i \in I_1, j \in J$ (9)

$P_{ij} \le (l_{ij} * a_{i(j-1)}) + (q_{ij} - l_{ij}) a_{ij} + BigM * (1 - E_{ij}) \quad \forall i \in I_2, j \in J$ (10)

$P_{ij} \ge (l_{ij} * a_{i(j-1)}) + (q_{ij} - l_{ij}) a_{ij} - BigM * (1 - E_{ij}) \quad \forall i \in I_2, j \in J$ (11)

Linear approximation constraints

$t_{ixy} \le b_{ix(y+1)} + BigM * (1 - Z_{ixy}) \quad \forall i \in I, y \in Y$ (12)

$t_{ixy} \ge b_{ixy} - BigM * (1 - Z_{ixy}) \quad \forall i \in I, y \in Y$ (13)

$t_{ixy} \le BigM * Z_{ixy} \quad \forall i \in I, y \in Y$ (14)

$\sum_y Z_{ixy} = 1 \quad \forall i \in I, x \in X$ (15)

$\sum_y t_{ixy} = T_i \quad \forall i \in I, x \in X$ (16)

$r_{ixy} = \beta_{ixy} + \alpha_{ixy} * t_{iy} \quad \forall i \in I, y \in Y, x \in X$ (17)

$f_{ixy} \le BigM * Z_{iy} \quad \forall i \in I, y \in Y, x \in X$ (18)

$f_{ixy} \le r_{ixy} + BigM * (1 - Z_{ixy}) \quad \forall i \in I, y \in Y, x \in X$ (19)

$f_{ixy} \ge r_{ixy} - BigM * (1 - Z_{ixy}) \quad \forall i \in I, y \in Y, x \in X$ (20)

$\sum_y f_{ixy} = F_{ix} \quad \forall i \in I, x \in X$ (21)

Variable formatting constraints

$E_{ij} \in \{0,1\} \quad \forall i \in I, j \in J$ (22)

$Z_{ixy}, t_{ixy}, r_{ixy}, f_{ixy} \in \{0,1\} \quad \forall i \in I, y \in Y, x \in X$ (23)

$T_i, Q_i, P_i > 0 \quad \forall i \in I$ (24)

$q_{ij}, p_{ij} \ge 0 \quad \forall i \in I, j \in J$ (25)

$F_{ix} > 0 \quad \forall i \in I, , x \in X$ (26)

Interpretation of the mathematical models

The key purpose of the model is to determine the value of review period T for each product. This variable variable has large impact on order quantity and number of orders placed per planning horizon, which play as decisive role in the value of cost function.

Objective function

The objective of the model is to minimize the result of cost function which includes four cost elements

Total purchasing cost: According to the periodic inventory review policy, the order quantity is the same among orders placed at the end of each period review. As a result, the total purchasing cost in planning horizon is product between purchasing cost per order and number of orders per planning horizon. The cost per order can be calculated based on the rule of discount for each items.

Fixed ordering cost: Based on the review period T, supply planner will need to place different number of orders for each item which will decide the total fixed ordering cost for that item in the whole planning horizon. The fixed ordering cost per order may include company's facilities and the maintenance cost of the computer system used to process that purchase order.

Holding cost: Inventory is an asset account that requires a large amount of cash outlay, and decisions about inventory spending can reduce the amount of cash available for other purposes. In this case, this type of opportunity cost is called holding cost which is defined two main factors: Average Inventory, Interest rate. In this problem, the author takes 0.0025/item/day as the interest rate while the average inventory, like many other inventory models, is combination between half of order quantiy and saftety stock.

Lost profit cost: When the standard normal distribution is used for estimating the market demand, we will face the risk of out of stock based on z-score extracted from chosen service level. To estimate the expected number of units loss, the standard normal loss function is applied

$$L(z_a) = \int_a^\infty (x - a)\varphi(x)dx$$

Where:

$L(z_a)$: Value of standard normal loss function at z = z_a

$\varphi(x)$: Standard normal distribution density function.

Then, the expected number of unit loss is defined as below:

$$U(z_a, \sigma) = L(z_a) * \sigma$$

Where:
σ: Standard deviation of random variable

Having got the expected number of units loss, we multiply this with profit per unit to get the lost profit cost per period.

Because the original objective function includes some non-linear factors, it will be approximate to **Linearized objective function** using the method discussed in the chapter of methodology.

Constraint

Market demand constraints:

(1) Review period should less than half of product's shelf life so that order quantity at the beginning of each review period should be sold before 50% shelf-life has passed.
(2) For each item, total order quantity is defined as the sum of order quantity at each discount level.
(3) For each item, total purchasing cost paid is defined as the sum of purchasing cost paid at each discount level.
(4) Total order quantity at the beginning of each review period should be able to fullfill the demand within review period.

Discount order quantity constraints:

(5)-(7) These two constraints define the discount level that can be applied based on the order quantity and given bounds of each discount level. If $l_{ij} \le Q_{ij} \le u_{ij}$, E_{ij}=1. Otherwise, E_{ij}=0.
(8)-(9) These constraints calculate the total purchasing cost for items which follow all-unit discount rule. If E_{ij} =1, then $p_{ij} = q_{ij} * a_{ij}$. Otherwise, there is no determined value for p_{ij}
(10)-(11) These constraints calculate the total purchasing cost for items which follow incremental discount rule. If E_{ij} =1, then $p_{ij} = (l_{ij} * a_{i(j-1)}) + (q_{ij} - l_{ij})a_{ij}$. Otherwise, there is no determined value for p_{ij}.

Linear approximation constraints:

General idea: *As discussed in previous part, the main idea for linear approximation is to separate the non-linear graph into many line segments created by pairs of points. Based on these pairs of points, the linear function of each line segments could be defined. In other words, the original non-linear graph is approximated by many linear functions and each of them will have a unique "applicable range" whose "applicable limits" is between a pair of points.*
Thus, the process to find the value of a non-linear function with the given input includes three steps. Firstly, the "applicable region" that the input belongs to should be identify. Based on the "applicable range", the answer for the question "Which linear function could be used to approximate the non-linear function?" shall be found. Finally, the approximating value will be calculated based on the value of input and linear function which is defined in previous step.

(12)-(14) These constraints are used to define which "applicable range" that the input belongs to. The result is shown by the binary variable Z_{ixy}. Z_{ixy} =1 if $b_{ixy} \le t_{ixy} \le b_{ix(y+1)}$ which means that input t_{ixy} belongs to "applicable range i". The applicable range i the pre-defined range which has lower limit b_{ixy} and upper limit $b_{ix(y+1)}$.

(15)-(16) These constrains mean that for all linear approximation process, there is only one "applicable range" whose linear function acrossing lower limit and upper limit, can be used to approximate the non-linear function. The input to this linear function is the decision variable T_i.

(17)-(21) These constraints simulate the linear approximation process. The value of original non-linear function with input t_{iy} can be approximated to the linear one whose slope-intercept form is: $r_{ixy} = \beta_{ixy} + \alpha_{ixy} * t_{iy}$. $\beta_{ixy}, \alpha_{ixy}$ are pre-determined parameters of for each applicable range of each linear approximation process. Constraint (18)-(20) emphasize that the approximation linear function could only be available if the input belongs to the applicable range with respect to that linear function. Otherwise, that linear function will show the 0 value. The constraint (21) is the summation for set of auxiliary variables including one non-zero value and many zero value. The result from constraint (21) is the final solution for non-linear approximation process.

Variable formatting constraints.

(22)-(26) Based on the use in mathematical model, the format of each decision will be setup with the right form so that the mathematical model could provide the applicable solution.

After formulating the problem to the mathematical model, this mathematical model will be converted to programming language so that the problem could be solved on a dedicated tool. The tool used that is used to solve this problem is CPLEX Studio IDE 12.8.0.

3. Business acumen in inventory planning:

As discuss in previous part, perishability risk of items in product portfolio can be categorized as below:

Shelf-life	Long	C1	C2	C3
	Medium	B1	B2	B3
	Short	A1	A2	A3
Perishability Risk		Slow	Normal	Fast
		Stock Movement		

Figure 11: Perishable Penalty convention

Based on the categorization about the perishable product as above, we classify these product groups into 3 risk groups:

- High risk: A1
- Medium risk: B1, A2, C1, B2, A3
- Low risk: C2, B3, C3

As classification for product groups as above, we will apply the different strategies to monitor each risk level to minimize risk and loss for the business while improve the cycle service level for the customers.

For high-risk products

For this kind of product which requires a lot of effort to manage well, the business should consider carefully about the profit to make decision whether they should continue to keep this product in the product portfolio that they distribute. If the profit is not high enough, the removal of this product shoul be taken into consideration. Otherwise, business could consider make-to-order strategy rather than make-to-stock strategy so as to maintain this product in the product categogy without pay much attention to track and mitigate the risk.

For medium risk products

For this group of products, the first solution is applying risk pooling techniques across products with high commonality in demand planning and supply planning so as to reduce the buffers that need to be prepared for demand uncertainties which are caused by demand variability. Relating to risk pooling across locations, Simchi-Levi in 2009 stated that: "Risk Pooling suggests that demand variability is reduced if one aggregates demand across locations. It becomes more likely that high demand from one customer will be offset by low demand from another, allowing a decrease in safety stock and therefore reduced average inventory". This philosophy can also be applied for risk pooling across products where demand from this product can be switched to another product with high commonality, thus decreasing the safety stock required for a pre-determined service level due to lower standard deviation in the demand. In the other words, in some case, single products with high commonality should be consolidated into group and demand forecasting process should be executed at product group level rather than single product level.

In addition, inventory planning team should have approriate plan to closely follow and make timely decision so that the risk is highlighted to relevant stakeholders and the collaboration among stakeholders shall be kicked off so as to mitigate the risk. This management can be conducted at the batch level instead of SKUs level because batch of product show the information about expiry date, and, managing at batch level will ease the effort paid for management. When there exists the risk aboout shelf-life, the inventory should make alarm to another department and find the solution for this risk. Relating this problem, there are two main departments that inventory teams will frequently contact with including sales department and warehouse department

Sales department

There is a fact that each type of customer will have different remaining shelf-life for the goods receipt. For example, for food product, the requirement from modern trade channels such as convenience store, supermarket, … etc is above 70% or 80% while traditional trade channels such as grocery or market... etc only require 50%-60% remaining shelf-life. Thus, goods allocation for order seems to be the popular solution from sales department. This solution is a win-win solution for inventory planning team and sales team. Due to optimal goods allocation based on requirement of specific customers and the actual condition of inventory, sales department can utilize all source of goods to minimize risk of out-of-stock.

The risk out-of-stock stated here is not the condition when there is no stock left at the warehouse, but the problem occuring when the remaining stock in the warehouse can not meet the requirement of the order from the perspective of remaining shelf-life. This problem is often due to the inapporiate goods allocation planning when receiving fresh stock – treat all order in the same way and allocate all stock with highest remaining shelf-life to all these orders without carefully consider the specific requirement of each order. For inventory team, approriate goods allocation for order will be helpful to minimize risk of goods perishability when all source of goods can have the good chance to move out the warehouse, thus increase the turnover rate and improve the overall performance of inventory health from the perspectives of not only supply chain, but finance.

Another solution from sales department shall be liquidation of perishable goods. Through this, creating special promotion plan for goods with short remaining shelf-life can help to prevent goods disposal and minimize loss of perishability risk. In case liquidation can not be executed, business can consider turning this amount of goods into FOC (free-of-charge) goods which is a strategy to boost the demand of other products so that the profitability is improved and the loss from perishbility risk can be partially covered.

Finally, in case the company have many warehouses at many places for inventory storage, stock allocation planning should be discussed carefully between sales department and inventory management department. In this plan, inventory management department is responsible for stock allocation decision making while sales department is the one which has the knowledge about the sales behavior and sales contribution of each area. Thus, with the information from sales team, inventory can make the optimal decision of warehouse-to-warehouse inventory transfer to transfer right product with the right quantity at right expiry date in right time and to the right place.

Warehouse department

Besides sales department, warehouse department is also a stakeholder that works closely with inventory management team. The solution from warehouse team is the use of approriate inventory picking rules.

Due to adding constraint of shelf-life, FEFO (First-Expired-First-Out) and LEFO (Last-Expired-First-Out) should be applied instead of simple FIFO (First-In-First-Out) and LIFO (Last-In-First-Out). In specific, while picking the products for order fullfillment, the FEFO is applied so that the product with higher risk could move out the warehouse as soon as before the expiry date. However, when picking activity serves process of warehouse-to-warehouse inventory transfers, the LEFO rules should be applied.

Especially, when the goods has short shelf-life in comparison with delivery lead-time, this rule should be followed strictly so as to lengthen the time that the amount of transfered goods become short shelf-life stock in the destination warehouse and can not be sold out.

For low-risk products

Low risk products are products having normal to fast movement rate while owning medium to long perishability risk. This group of products requires business little effort for management. However, the sales behavior and inventory condition should also be tracked so that the risk is highlighted timely when there is any unfrequently pattern appearing. For example, when the sale patterns should be reviewed frequently, the undesirable drop in market demand will be detected immediately and adequate actions will be taken timely to prevent low risk amount of inventory from transforming to medium or high risk one.

4. Solution Analysis

4.1. Demand Planning

To have more insightful information about the result of the forecasting model, an analysis at the product category is performed.

As discussed in previous, MAPE which stands for Mean Absolute Percentage Error is the first metric to be used for evaluating the forecasting result. Let “High Performing” denotes forecasting result with MAPE < 20%, “Medium Performing” for ones with MAPE in the interval [20%, 40%] and “Low Performing” for remaining results, we have tables of frequencies as below:

Table 6: Frequency table to evaluate forecast accuracy performance with respect to perishability groups

Group	High	Medium	Low
A1	1	1	0
A2	1	1	0
A3	2	0	2
B1	0	4	1
B2	3	2	2
B3	7	3	3
C1	0	2	1
C2	0	0	1
C3	2	3	0

Table 7: Frequency table to evaluate forecast accuracy performance with respect to profit groups

Group	High	Medium	Low
A	7	3	8
B	4	7	1
C	5	6	1

For better comparision and conclusion, we transform the frequency table to probability table and observe the result.

Table 8: Probability table to evaluate forecast accuracy performance with respect to perishability groups

Group	High	Medium	Low
A1	50%	50%	0%
A2	50%	50%	0%
A3	50%	0%	50%
B1	0%	80%	20%
B2	43%	29%	29%
B3	54%	23%	23%
C1	0%	67%	33%
C2	0%	0%	100%
C3	40%	60%	0%

Table 9: Probability table to evaluate forecast accuracy performance with respect to profit groups

Group	High	Medium	Low
A	39%	17%	44%
B	33%	58%	8%
C	42%	50%	8%

From the perspective of perishability risk management, the forecasting model provides the "Medium performing" to "High performing result" for majority of items with high perishability risk (Group A1 to A3). This performance can be stated to be maintained for product with medium risk (B1 to B3) in spite of having increase in low-performing results. Finally, the product group with lowest perishability risk (C1 to C3) has the most risky results due to having majority of "Low performing" to "Medium performing" result and can be considered as the chance for improvement.

From the perspective of profit class or lost profit risk, there exists the degeneration in the results for item with profit class A because of the higher proportion in low-performing result. However, the performance for items that needs paying low to medium attention has been significantly better with the low proportion of "Low performing forecasting result".

In conclusion, the analysis on MAPE for each product has provided an overall concept about the performance of forecasting model. Through this review, some competitive advantages and potential risk has been highlighted which can be utilize to evaluate the model.

However, to have more insight for impovement suggestion, more deep-dive performance metric should be used in order to discover key characteristics of model's performance that behind the result. Especially when making discussion about lost-profit risk or perishability risk, concern of people in supply chain is not only the error of the forecast result, but also whether the direction of that error is positive or negative. In other words, they need to know that the current prediction value is higher or lower than the actual sales – which can be defined as "under-forecast" or "over-forecast". For this case, "Forecast Bias" (FB) can show its great capability.

Let "Under-forecast" denotes forecasting results with negative bias (prediction value < actual value) and "Over-forecast" for the ones with positive bias (prediction value > actual value), we come up with the tables of frequency as below:

Table 10: Frequency table to evaluate forecast bias with respect to perishability groups

Group	Over-forecast	Under-forecast
A1	0	2
A2	1	1
A3	1	3
B1	0	5
B2	1	6
B3	3	10
C1	0	3
C2	0	1
C3	1	4

Table 11: Frequency table to evaluate forecast bias with respect to profit groups

Group	Over-forecast	Under-forecast
A	4	14
B	1	11
C	2	10

Similar to previous analysis, we need to transform the frequency table to probability table so that more informative insight could be revealed.

Table 12: Probability table to evaluate forecast bias with respect to perishability groups

Group	Over-forecast	Under-forecast
A1	0%	100%
A2	50%	50%
A3	25%	75%
B1	0%	100%
B2	14%	86%
B3	23%	77%
C1	0%	100%
C2	0%	100%
C3	20%	80%

Table 13: Probability table to evaluate forecast bias with respect to profit groups

Group	Over-forecast	Under-forecast
A	22%	78%
B	8%	92%
C	17%	83%

From the perspective of perishability risk management, the forecast bias results show the little risk of perishability for all product group because they show the tendency of under-forecast result for most of products in porfolio. In other words, the forecasting model is likely to provide the prediction value that is lower than actual sales. In conclusion, this model seems to perform well in the objective of minimize perishability risk.

However, perishability risk and profit risk is two different stories with opposite target. While "under-forecast" is good news for who pay much attention to perishability risk, this is a big concern for who want to increase the profit. This concern is called "out-of-stock". From the perspective of profit class or lost profit risk, there exists the high risk of out-of-stock for all product groups due to the high proportion of under-forecast results.

Within the scope of this research where salvage activities is not considered for perishable products and according to the case study where the company, as many new players in the market, want to maintain low operation cost before expanding the market, this solution seems to be the safe one because in this situation, the cost paid for over-forecast (cost of goods sold) is higher than the cost paid for under-forecast (profit of goods sold). However, high possibility of under-forecast is a problem needing to be solve if company want to increase the cycle service level, expand the market and earn more profit in the future.
While having unbalancing performances for two objectives – "too good to minimize perishability risk " and "need much improvement to minimize lost profit risk", there exists the big opportunities to balancing two objectives so as to improve overall total cost.
For this problem, there is some suggestions that can be considered including:

- Adjust loss function for training process, adding the factor relating to direction of error in addition to the amount of error.
- Increasing safety stock level (usually through service level) for items with high tendency of under-forecast result while having high profit.

4.2. Supply Planning

The result from the model could help to define the optimal value objective function with parameters of the inventory periodic review policy including review period as well as corresponding purchase quantity and level of discount.

Sensitity Analysis

As discussed in previous parts, sensitivity analysis is the process of making small adjustment in some types of parameters and observing the impact on the solution. Based on this observation, conclusion for some insightful information about the model shall be drawn and some suggestion will be given for further improvement.

There are two parameters which will join in sensitivity analysis process including Remaining shelf-life requirement and Planning bucket

Remaining shelf-life requirement

While discount quantity policies encourage supply planners to purchase as much as possible so that they benefit the opportunity to cut down kinds of purchasing cost, the remaining shelf-life (RSL) requirement seems to be a big constraint that prevent supply planners to place big order and hold such a large amount of inventory. Perishability is always a big concern with short shelf-life products. For food products, this concern shall even be more important, and customers always want to maximize the remaining shelf-life when they purchase the product.

However, when the company expand the market, build its distribution network with variety of customers whose requirement is different from each others, the company can be more flexible in making purchasing decision. For example, while some modern trade channels such as supermarket, shopping malls, have high requirements for remaining shelf-life (about higher than 80% RSL), some traditional trade channels like markets only require their distributor to supply goods with higher 40% RSL. Hence, the constraint $1 \leq T_i \leq 0.2n_i$ in the mathematical model can be adjusted into $1 \leq T_i \leq 0.6n_i$.

In a common sense, this change can create chance for bigger solution in the review period because the more goods can be purchased which can save cost per unit purchased according to discount policy.

Planning bucket

Planning bucket refers to a unit of planning period such as daily, weekly, bi-weekly, monthly, or even quarterly. In other words, the review period in our problem should be the multiple of the planning bucket when both is at the same unit. For example, if the planning bucket is weekly, the review period in days should be the multiple of seven. The current planning bucket used in this research is daily.

To analyze the effect planning bucket to the result, the planning bucket constraint is adding to model:

T mod p = 0

Where:

T: Review period in days

mod: Modulo operation

p: Planning bucket in days

For this sensitivity analysis process, we consider four types of planning bucket: daily, weekly, bi-weekly, monthly and nine level of remaining shelf-life ranging from 0.3 to 0.7 with step 0.05. Based on this, a comparison about optimality among types planning bucket and levels of remaining shelf-life shall be conducted . In addition, the effect of adding planning bucket constraint on relationship between remaing shelf-life and total cost will be performed through this sensitivity analysis process. Finally, the recommendation for negotiating suitable standard remaining shelf-life will be given so as to provide a win-win solution for both sides.

Experiment description:

As mentioned in previous part, there are two main parameters that will be adjusted in this sensitivity analysis process as below:

Remaining shelf-life	Planning buckets
0.3	Daily
0.35	Weekly
0.4	Bi-weekly
0.45	Monthly
0.5	
0.55	
0.6	
0.65	
0.7	

Table 14: Experiment description for sensitivity analysis process

Thus, for each type of discount policy, we will have totally 36 running cases for each item. Through each running case, the result of review period and total cost will be observed and visualized by data table and approriate types of charts so that the insightful information could be extracted and conclusion about the relationship between independent variables and dependent one, and interaction between parameters will be drawn.

Legends

To visualize the discount levels applied, the below color gradient legends is used

	Discount level 1
	Discount level 2
	Discount level 3
	Discount level 4

Figure 12: Legend visualizing discount level on sensitivity analysis table

All-unit discount policy

Table 15: Sensitivity analysis for all-unit discount items using monthly planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1		60	60	60	90	90	90	90	120
T2		60	60	60	90	90	90	90	120
T3		60	60	60	90	90	90	90	90
T4		30	30	30	30	30	30	30	60
Z	0	44,473	44,473	44,473	42,743	42,743	42,743	42,743	39,360

Table 16: Sensitivity analysis for all-unit discount items using bi-weekly planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1	42	56	70	70	84	98	98	112	112
T2	42	56	70	70	84	98	98	112	112
T3	42	56	70	70	84	84	84	84	84
T4	14	28	28	28	42	42	42	56	56
Z	45,209	44,712	44,476	44,476	42,723	41,793	41,793	39,357	39,357

Table 17: Sensitivity analysis for all-unit discount items using weekly planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1	49	56	63	77	84	98	105	105	105
T2	49	56	70	70	84	98	105	105	105
T3	49	56	70	77	84	84	84	84	84
T4	21	28	35	35	42	49	49	56	56
Z	45,145	44,712	44,461	43,489	42,723	41,125	40,318	39,354	39,354

Table 18: Sensitivity analysis for all-unit discount items using daily planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1	54	58	72	80	80	99	103	103	103
T2	54	62	72	81	83	99	104	104	125
T3	54	62	72	81	86	86	86	86	84
T4	27	31	36	40	45	49	54	58	65
Z	45,118	44,470	44,450	42,873	42,720	40,958	39,354	39,352	39,235

Incremental discount policy

Table 19: Sensitivity analysis for incremental discount items using monthly planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1		60	60	60	60	60	60	60	60
T2		60	60	60	90	90	90	90	90
T3		60	60	60	90	90	90	90	90
T4		30	30	30	30	30	30	30	30
Z	0	29,447	29,447	29,447	26,368	26,368	26,368	26,368	26,368

Table 20: Sensitivity analysis for incremental discount items using bi-weekly planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1	42	42	70	70	70	70	70	70	70
T2	42	56	56	56	84	98	98	98	98
T3	42	56	56	56	84	98	98	98	98
T4	14	28	28	28	28	28	28	28	28
Z	37,696	30,832	30,149	30,149	27,477	24,802	24,802	24,802	24,802

Table 21: Sensitivity analysis for incremental discount items using weekly planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1	42	42	70	77	77	77	77	77	77
T2	49	56	56	77	84	98	105	105	105
T3	49	56	56	49	84	98	105	105	105
T4	21	28	35	35	35	35	35	35	35
Z	34,307	30,832	29,145	27,885	26,216	23,541	22,472	22,472	22,472

Table 22: Sensitivity analysis for incremental discount items using daily planning bucket

Remaining shelf-life	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
T1	46	62	72	77	70	77	77	77	77
T2	54	60	60	81	90	99	106	106	106
T3	53	53	53	81	90	99	108	108	108
T4	27	31	36	37	37	37	37	37	37
Z	31,177	29,188	28,114	26,693	24,751	23,163	21,996	21,996	21,996

Interpretation of table result

In general, for all planning bucket, decreasing the remaing shelf-life requirement will never increase the total cost - it will decrease or remain unchanged. This can be explained that the ease of shelf-life requirement will let the business to own stock longer. In other words, the business can buy more inventory at a time which will take much more time to sells. As a result, this fact will encourage business to place bigger order to enjoy the cost benefit including benefit in purchasing cost and ordering cost from discount policies. Especially, this correlation will even become more sensitive if the rate of changing in holding cost - directly proportional to time of keeping stock on hand, is smaller than rate of changing in purchasing cost and ordering cost. This is because purchasing cost and ordering cost have larger impact on the total cost than holding cost in this situation.

Observing impact of planning bucket, it can be seen from the result table that for all remaining shelf-life level, adding planning bucket constraint shall decrease the optimal value of objective function. This can be proved through the total cost comparison between bi-weekly planning bucket and weekly planning bucket, or between the daily planning bucket and the others planning bucket. In addition, this constrait will require that the remaining shelf-life standard should be lower so that the optimal size of order can become bigger and reach higher discount level. For example, in the part of all-unit discount policy, using daily planning bucket, the optimal order quantity item 3 will reach the discount level 4 when remaining shelf-life standard is 0.55. However, when using monthly or bi-weekly planning bucket, this item will not benefit discount level 4 until remaining shelf-life standard is decreasing to 0.5. Another example item 1 in the part of incremental discount policy. Using the daily planning bucket, the optimal order quantity of this item reach the discount level 3 when the remaining shelf-life standard is up to 0.65. However, the required remaining shelf-life is 0.6 if the weekly or bi-weekly planning bucket is used. The impact of planning bucket can be explained by the constraint put on the review period. In specific, the review period must be the multiple of 7 for weekly planning bucket, 14 for bi-weekly planning bucket and 30 for monthly planning bucket. Thus, adding planning bucket constraint will make the model return the result which is far from the "true" optimal point. In other word, for this situation, due to the effect of planning bucket constraint, the review period will be rounded instead of freely move to the actual optimal point, thus reduce the ability to achieve higher điscount level. This is the reason why the remaining shelf-life should be đeducted if the business want to increasing the order quantity for better discount level. Applying this interaction between remaining shelf-life and planning bucket, business can make a balancing decision so as to minimize the total inventory cost while providing better service for service level for its customer.

Chart visualization for results

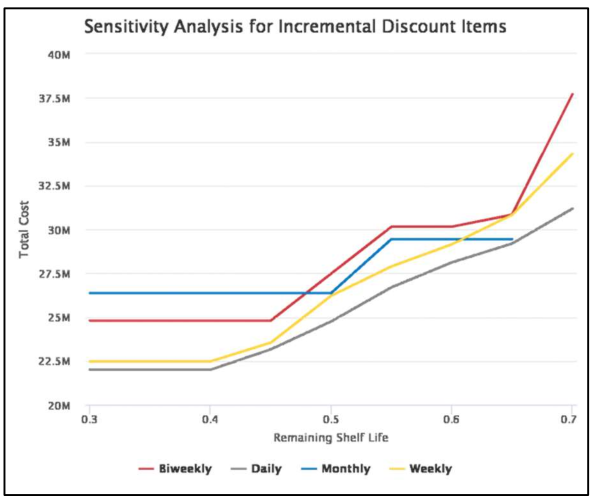


Figure 13: Sensitivity Analysis for Incremental Discount Items

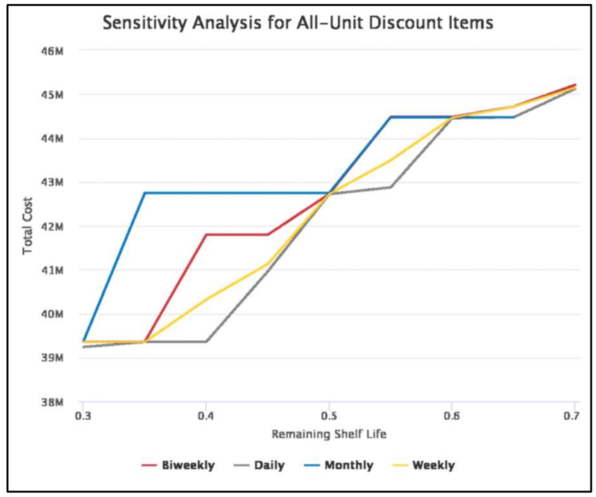


Figure 14: Sensitivity Analysis for All-Unit Discount Items

It can be seen from the line chart that this visualization once again emphasized the conclusion stated in the previous part about the impact of remaining shelf-life standard and planning bucket towards the value of total cost. In addition, above graphs reveal different impact levels of remaining shelf-life standard towards the total cost with respect to each planning bucket. In general, the sensitivity of total cost towards the change of remaining shelf-life standard become smaller for big planning bucket. For example, in the part of incremental discount, for monthly planning bucket, the result of total cost remains unchange despite of increasing the remaining shelf-life standard from 0.3 to 0.5. However, for daily or weekly planning bucket, this result is only kept stable when raising the remaining shelf-life standard from 0.3 to 0.4, and make a lift if this parameter continues to be increasing.

This is on the contrary to the common sense where increase service level always increase cost for the business. Although there exists less closely relationship between remaining shelf-life standard and total unit cost for big planning bucket, when keep increasing this parameter to some "critical" point, it is more likely to appear big lifting step for the value of total cost in big planning bucket rather than smaller planning bucket. The explanation above behaviors is still the "rounding" story. Firstly, when the change in material requirement that made by adjustment in remaing shelf-life standard is not big enough to satisfy the "divisibility" requirement in planning bucket constraint, especially for big planning bucket constraint, the outcome of optimal solution will not change. Hence, the optimal value of objective function will remain unchanged. Secondly, when the increasing the remaining shelf-life standard to some "critical" point where the accumulative change in material requirement can satisfy the "divisibility" requirement in planning bucket constraint, it will appear an increase in the optimal value of review period, thus, rise the total cost. This is the reason why it is more likely to appear "big lift" in big planning bucket rather than the smaller one.

Applying above insightful information, supply planners can make better decision in negotiating with customers by providing the best service level while not affecting much to the profitability of the business by using the sensitivity analysis of parameters to the result of objective function and the trade-off strategy between the remaning shelf-life standard and planning bucket. In detail, they can when they met the constraint from planning bucket, adjust remaining shelf-life can help them to reduce the loss. In addition, observing the behavior of the total cost function in big planning bucket, they can choose to profit highest service level while do not affect profitability of company when negotiating and persuade their customers.

5. Conclusion

In this study, we presented a forecasting model in terms of Gradient Boosting method. Thanks to the feature engineering and model tuning process in model training, the final model provides us the average forecast accuracy is up to 75% where the best performance is up to 91%. Further analysis shows that majority of the forecasting results fall in the group of "Medium Performing" to "High Performing" for products with medium to high risk or perishability. This performance is still maintained when the analysis performed in the aspect of profitability despite of having some degeneration. Evaluating the result by forecast bias, this model mostly provided under-forecast results which can prevent the risk of perishability risk for short shelf-life product.

In addition, we also applied linear programming to solve the optimization problem relating to design the optimal inventory ordering policy with consideration of discount quantity and shelf-life requirement constraint. While there is some non-linear function in the model, we have used approximation technique that transform the non-linear graph to a set of line segments determined by pairs of point. After getting the solution, we continue to seek for more insightful information by conduct the sensitivity analysis process with the participation of planning bucket and remaining shelf-life. Through this research, many relationships between remain shelf-life, planning bucket and total inventory cost have been discovered and created trustworthy reference for business when they have to negotiate with customers for win-win agreement for both sides.

Finally, some solutions from the perpectives of business were given in additon to engineering solution. According to this suggestion, inventory team should consider applying some risk pooling technique in demand and supply planning process to minimize risk relating to profitablity and perishability of goods. Besides, the close collaboration between departments is also required so that each department can achieve its goal and hence, the overall performance of the organization is improved.

Future research directions in this study include (a) considering larger datasets and a more variety of AI algorithms techniques to improve forecasting model (b) adjusting the loss function to balance the risk of perishability and lost profit (c) design inventory ordering policy, considering variety of remaining shelf-life requirement from customes.

Reference

Vlckova, Vladimira, and Michal Patak. (2010). "Role of Demand Planning in Business Process Management." *The 6th International Scientific Conference "Business and Management 2010". Selected Papers*, 2010, doi:10.3846/bm.2010.151

Stadtler H. (2005). Supply chain management and advanced planning—basics, overview, and challenges. *European Journal of Operational Research*. 2005;163(3):575-588. doi: 10.1016/j.ejor.2004.03.001

Sodhi P, Awasthi N, Sharma V. (2019). Introduction to Machine Learning and Its Basic Application in Python. *SSRN Electronic Journal*. Published online 2019. doi:10.2139/ssrn.3323796

Chen S. (2021). A Basic Machine Learning Workflow in Production. Medium. Published June 22, 2017. Accessed January 26, 2021. https://medium.com/eliza-effect/how-machines-learn-d9e9a3e6f97c

Chen FL, Ou TY. (2011). Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*. 2011;38(3):1336-1345. doi: 10.1016/j.eswa.2010.07.014

Lee AHI, Kang H-Y, Lai C-M, Hong W-Y. (2013). An integrated model for lot sizing with supplier selection and quantity discounts. *Applied Mathematical Modelling*. 2013;37(7):4733-4746. doi: 10.1016/j.apm.2012.09.056

Esfandiari N, Seifbarghy M. (2012). Modeling a stochastic multi-objective supplier quota allocation problem with price-dependent ordering. *Applied Mathematical Modelling*. 2013;37(8):5790-5800. doi: 10.1016/j.apm.2012.11.011

HEALTHY URBAN PLANNING APPLICATION IN HO CHI MINH CITY, A POTENTIAL FOR FUTURE DEVELOPMENT TOWARD SUSTAINABILITY

Vo Minh Phuc [1*], Vo Trong Cang [2]

[1] *IMALOG, Binh Duong University, Binh Duong, Viet Nam, vmphuc@bdu.edu.vn*

[2] *IMALOG, Binh Duong University, Binh Duong, Viet Nam, vtcang@bdu.edu.vn*

Abstract: Healthy Urban Planning (HUP) is an advanced urban planning approach promoting health factors' role in the existing urban planning paradigm, which is also a potential solution for current Ho Chi Minh City's (HCMC) problems. Recently, HCMC is facing several negative problems, as consequences of rapid urbanization decades, which are threating inhabitants' life, such as: pollution and related diseases, traffic congestions and accidents, urban sanitation, and flooding, etc. Aiming to promote HUP and its principles, this article introduces HUP approach and Healthy Cities concept. Next, the recognition of health factor in sustainable urban development (SUD) goals are discussed through WHO publication and agenda. Next, problems and related adverse health impacts, relating to urbanization, are reviewed to highlight the current problems that HCMC inhabitants are struggled with. Finally, this research focuses on qualitative analysis on the potential of HUP application in term of problem solving and HCMC advantages in HUP integration into current sustainable development planning paradigm.

Keywords: Healthy Urban Planning (HUP); Sustainable urban development (SUD); Ho Chi Minh city (HCMC)

1. Introduction

1.1. Healthy Cities and Healthy Urban Planning concepts

A Healthy City is "one that continually creates and improves its physical and social environments and expands the community resources that enable people to mutually support each other in performing all the functions of life and developing to their maximum potential" (WHO, 2021a). The Healthy Cities approach is a city development method that puts health high on the political and social agenda of cities and to build a strong system for public health care at the local level. In addition, a Healthy City strongly emphasizes equity, participatory governance and solidarity, inter-sectoral collaboration and action to improve the health condition of its inhabitants. (WHO, 2021a)

Healthy Urban Planning (HUP) is planning concept that puts the needs of people and communities at the heart of the urban planning process and considers the implicates of decisions for human health and well-being (WHO, 2021b).

The similarity between healthy urban planning and sustainable urban planning is the key to accomplish the planning objectives is finding balances between social, environmental and economic pressures.

The healthy urban planning and healthy city have been developed and applied since 1986, as a result of Ottawa Charter for Health Promotion (WHO, 1986).

1.2. The recognition of health factor in sustainable urban development goals

Since the publication of Health of All also called "Health21", the public health has been recognized as one of major keys for sustainable development achievements. In SUD 2030 Agenda (UNO, 2018), 'Good Health' has a central place in Sustainable-Development-Goals 3, "Ensure healthy lives and promote well-being for all at all ages", underpinned by 13 targets and 17 goals that cover a wide spectrum of WHO's work. (Figure 1)

In the context of this research, HUP and Healthy City are introduced at beginning stages. Next, practical public health problems of Ho Chi Minh City (HCMC) are identified. Last but not least, HUP approach adaptivity potential in situation improvement is discussed.

Figure 1. 2030 Agenda: 17 goals for saving the planet (United Nation Organization, 2018)

2. Research methodology

Due to objective and scale of the current paper is forming a literature-base study and fundamental knowledgebase for further study.
Therefore, the major applied methodology is literature review data collection method. Moreover, in order to form the statements, the gathered data from literature review is sorted and compared analysis internally.

In details, the literature review includes literature concepts, legislation and scientific relationship between health risk exposure, health effects and urbanization-related issues; practical data about Ho Chi Minh City problems through scientific report, high-legit newspaper, such as: fine particles and chemicals pollutions; injuries due to traffic, traffic congestion; urban sanitation and flooding problem. Next, the input data is sorted, compared and summarized to form completed input data. Finally, the qualitative analysis is conducted on these input data; as the results, the conclusion on the HUP adaptability into current Sustainable Urban Planning approach of HCMC.

3. Healthy Urban Planning (HUP) In Modern Urbanization

As many adverse health effects from industrial urbanization during 18th Century, the urban planning has evolved continuously. As results, several urban planning schools are established and develops their own concept and system, notably there is the concept of sustainability and Sustainable Urban Development (Duhl and Sanchez, 1999).

Based on urban sustainability theory, The New Urbanism or Neo-traditional Planning is emerged. This primary design ideas of this New Urbanism are to form pedestrian-centered neighborhood where most of social and economic amenities are located within 5-minutes-walking range and inhabitants' settlement is planned surrounding public transportation hub and set up mix-used land blocks instead of single land-use areas, like an old planning paradigm (Duhl and Sanchez, 1999).

The new urban planning paradigm has improved living quality since it encourages positive and active lifestyle as well as enhance economic activities through shifting the urban land use planning, from mobility planning to accessibility planning, limiting private motorization demand while encouraging non-motorized transport and public transport enhancing.

As the development of urban study and social recognition on the importance of health, HUP is emerged and promoted by WHO, first in European Region, now throughout the world; currently, the Alliance for Healthy Cities (AHC) are formed in 2003 to connect Healthy Cities together (AHC, 2021a). Healthy Cities integrated principles of HUP into their planning system, which aim to cope the adverse effects of an urban environment over health. Up till 2021, there are 9 Full members Nations joined in the Alliance, with more than 60 participated cities (AHC, 2021b), including Hue, the tourism and cultural city in the central of Vietnam.

In the next section, related health problems occurring in HCMC is identified: air pollution and respiratory diseases, traffic accident and congestion, sanitation and urban flooding. Last but least, the HUP approach for improvement is proposed as a potential solution.

4. Ho Chi Minh Urbanization – Related Health Problems

For decades of rapid urbanizing, Ho Chi Minh City (HCMC) is now facing many problems, such as: environmental pollution, traffic accidents, congestion, over exploitation of natural resources… Particularly, transport activity is considered as the major source causing aggressively adverse health impacts. (Phuc and Cang, 2021)

4.1. Air pollution and respiratory diseases

Air pollution is one of the most notable problems in HCMC recently (Figure 2). First, the density of PM2.5 particle recorded as reached 40μg/m^3 (2017), 4 times higher than WHO suggestion at 10μg/m^3. The air pollution leads to the spread of respiratory diseases in community, especially youth, over 90% of less-than-5-years-old children in HCMC were infected to respiratory disease (2012) (Bang et al., 2017).

Figure 2. Air Pollution in areas HCMC in 2019.

The reason for the condensed pollutant fog is combination of low-temperature weather and the high amount of emission sources, such as: transport vehicles, industrial and daily life activities. Amongst many sources, transport activities emit many emissions overall, especially CO, NOx and NMVOC (figure 3) (Bang et al., 2017).

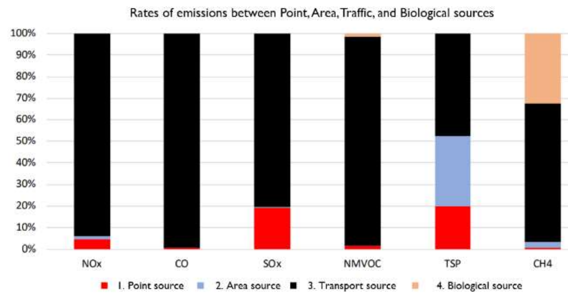


Figure 3. Rates of emissions between Point, Area, Traffic and Biological sources of emissions. (Bang et al., 2017)

4.2. Traffic accident and congestion

Another public health threats of HCMC are traffic congestion and accident.

Firstly, the traffic congestion harms health greatly, since it multiplies the vehicle pollution rate through multiple interactions (Le et al., 2012). Basically, the traffic congestion lowers the travel speed which mean increase travel duration and human exposure rate to pollution. Furthermore, congestion condenses vehicle – related pollutants alongside the road since the density of the pollutant gas depends on the movement of the traffic flow, or it can be said that the speed of vehicles traveling on the road (Zhang and Batterman, 2013). Thus, the congestion area is crowded with vehicles; therefore, the pollution at the intersection will rise dramatically. Finally, driving patterns are change during congested time, which leads to an increase in number of speedups, slowdowns, stops and starts, which increase emissions compared to "cruise" conditions, especially with high power acceleration engines (Benson, 1989).

In case of HCMC, the congestion occurs throughout the city, especially at main transport route during rush hours. In 2009, there are more than 61 30-minutes-or-longer congested point, especially 12 40-minutes-or-longer cases[1]. The result recorded in 2009 is higher than the 200 recorded result, 23 cases increased. During time, the traffic congestion has improved continuously due to the efforts from municipality. In 2016, the city had 37 points of traffic congestion, in 2018 it decreased to 28 points and in 2019, the number dropped to 22 points of traffic congestion (Tien Luc, 2020).

Secondly, the traffic accident is the next notable traffic – related problem, which causes 315,000 deaths in 2010 and 316,000 in 2013 due to traffic accident in South – East Asia region (JICA, 2008). In Vietnam, traffic accident and injury are concerned as a severe problem, which must be solved aggressively to archive the sustainable development. In 2007, the statistical survey, conducted by JICA (UNESCAP, 2014), shows that there are 14,727 traffic accidents cases which lead to 12,757 deaths and 11,288 injured people. More details, fatality ratio and injury ratio (per 10,000 residents and 10,000 vehicles) are still at high number when comparing with the motorization rate of Vietnam during the same period. In positive side, the record shows the huge improvement compared to the previous period.

In case of HCMC and Ha Noi, the number of accident cases recorded is much higher than the others city in Vietnam. Therefore, HCMC government has implemented many solutions aiming to improve the situation. Noticeably, the comprehensive solutions have been implemented throughout the city and on many development sectors. As result, in 2019, the city recorded 3,427 accidents, killing 641 people, and injuring 2,406 people; compared to 2018, it reduced 213 cases, decreased 74 deaths and decreased 69 people injured. (Tien Luc, 2020), (Tu Giang, 2020)

4.3 Urban sanitation and urban flooding

In Viet Nam, the urban sanitation and urban flooding are degrading living environment, in Southern region especially. First, in Viet Nam, less than 10% urban wastewater is treated centrally by public infrastructure (ADB, 2015). The most of urban household-used wastewater is treated using on - site sanitation, such as septic tanks. As results, contamination of groundwater and surface water is inevitable (Yen and Anh, 2021). Facing the sanitation problem, Viet Nam government, especially HCMC municipality, has cooperated with international organizations for implementing wastewater treatment - environmental improvement projects in the sustainable development manner, such as Asian Development Bank, World Bank. Particularly, up to June 2020, Viet Nam government has cooperated to conduct 39 projects, on sanitation and water supply sectors, with World Bank Organization since 2013 (WB, 2021). One of the most well-known projects is Nhieu Loc – Thi Nghe Basin Project in 2001 for its success and positive effects.

1 Mr. Tran Quang Phuong, Director of Ho Chi Minh City Department of Transport, stated at the meeting to discuss solutions to prevent traffic jams in Ho Chi Minh City in October 2009.

Next, urban flooding, because of climate change, has strongly affected HCMC inhabitants' living environment. In detailed, flooding affects HCMC's environment, public health, economy, and traffic aspects. During flood, the urban wastes are drifted by flood water, leading to water pollution and many diseases such as skin allergy, cholera, dysentery, especially dengue fever. The level of flooding has worsened through years despite of efforts from HCMC municipalities; detailed, municipalities pay VND7 trillion ($304 million) for the 2016-2020 period to fight urban flooding. Since 2016, HCMC has implemented the VND10-trillion anti-flooding project, which is the hope against this climate change phenomenon. Finally, the urban flooding will get worsened in the future if the implementing solutions are not worked as expected, due to the increase of climate change effects (Ha An, 2020), (Mook et al., 2016) and (Ha Mai, 2021).

5. Healthy Urban Planning for Sustainable Ho Chi Minh City

The potential for healthy urban planning application in HCMC is evaluated using qualitative analysis method on different sectors.

First, HCMC development orientation, aiming to achieve sustainability and improve quality of life, is the fundamental advantage for HCMC to adapt HUP into its current planning system. Moreover, the similarities between SUD and HUP are the huge advantages for HCMC municipality to adopt profitable HUP principles into its SUD mechanism. Some notable similarities are: (1) the human-centric development that provide the synchronized policies and infrastructure that can save up resources and maximize its use; (2) shared resources for SUD could be used for improving public health and social welfare system. Finally, the process reaching healthy state could help HCMC municipality achieve its sustainable development goals (WHO, 2021a and 2021b).

Second, the healthy urban planning concept and its principles are the perfect fill to improve the holes of current development of HCMC in term of solving current problems. The new planning would be more comprehensive than it used to be by putting health factor at the center for decision-making process, which used to be considered as a separated or supplement factor of traditional urban planning procedure. The Healthy Urban Planning binds the key health objectives into all of development actions of current city area that will make the decision-making calculation change aggressively (UNO, 2018). The changes have multiple positive impacts on City quality of life: (1) economic and social enhancement by reducing payment for illnesses cure and health care, increasing happiness and wellbeing which increase social productivity, etc.; (2) improve sustainability of living environment as well as urban environmental quality which are the fundamental for good health and wellbeing; (3) emissions reduction overall which decrease the pollution-related diseases and slowdown climate change effects, etc.

Finally, the Healthy Cities and HUP concept are not completely new to Viet Nam government, that would be an advantage for HCMC municipality to adopt HUP principles to current development mechanism through learning current case study. More details, Hue City, a culture and tourism hub in the middle region of Viet Nam, has participated in AFHC network since 2003. Afterward, local government and AFHC's experts cooperated to implement some successful healthy development projects: healthy market at Thong market, healthy public office, healthy community (Phu Mau settlement), etc. (AHC, 2014).

6. Conclusions and Recommendations

6.1 Conclusions

HCMC is struggling with problems from its rapid urbanization and City inhabitants is facing severe health problems days by days. Environmental pollution, pollution-related diseases, traffic accidents, congestions and flooding are identified as main threats for public health and City quality of life. The current urban planning mechanism, which based on SUD concepts and principles, is the right direction but unfortunately is not enough to solve existing health problems. As results, HUP appears as a perfect fit for the current situation.

HUP places public health factor at the center of decision-making, providing positive solutions forward Healthy state. Therefore, urban development projects based on HUP principles will improve the health problems of HCMC directly. Fortunately, HCMC has some advantages in adopting HUP principles into its current planning mechanism: (1) the similarity in City development orientation as well as similarities between SUD and HUP concepts; (2) HCMC facing problems can be solved by implementing Healthy projects; and (3) the existing case studies provides knowledgebase and practice experience for HUP adaptivity in HCMC context.

6.2 Recommendations

This scientific article is the first step that provide the knowledgebase in the context of HUP adaptability to urbanizing areas. Therefore, to conduct the comprehensive knowledgebase about HUP topic and its implementation into Vietnamese cities, more and 'further' studies are required in the future.

Within the limited scale of this article, the research describes very briefly current problems of HCMC as well as its advantages in term of HUP implementation into practice. The future research and scientific articles with more practical experience and statistical data would be a great contribution to current knowledgebase of HUP, SUD and urban development in general.
Some future scientific topics for our topics are:
(1) Relationship between inhabitants' health effects and HCMC's urban form, as well as transport sectors.
(2) Health development integration into Vietnamese urban development planning and legislation system.
(3) Modification of conventional urban transport into a health-oriented transport system.

References

WHO (2021a). What is a healthy city? Retrieved date: March 22nd, 2021a, from WHO: Regional Office for Europe, https://www.euro.who.int/ru/health-topics/environment-and-health/urban-health/who-european-healthy-cities-network/what-is-a-healthy-city.

WHO (2021b). Healthy urban design. Retrieved date: March 22nd, 2021b, from WHO: Regional Office for Europe, https://www.euro.who.int/en/health-topics/environment-and-health/urban-health/activities/healthy-urban-design.

WHO Regional Office for Europe (1986). Ottawa Charter for Health Promotion. WHO Regional Office for Europe, Copenhagen.

UNO - United Nation Organization (2018). Sustainable Development Goals. Retrieved date July 6th, 2021, United Nations: Department of Economic and Social Affairs, https://sdgs.un.org/goals.

L.J. Duhl and A.K. Sanchez (1999). Healthy Cities and The City Planning Process – A Background Document on Links Between Health and Urban Planning. WHO Regional Office for Europe.

Alliance for Healthy Cities (AHC, 2021a). About The Alliance. Retrieved date July 6th, 2021, Alliance for Healthy Cities: http://www.alliance-healthycities.com/htmls/about/index_about.html

Alliance for Healthy Cities. Members (AHC, 2021b). Retrieved date July 6th, 2021, Alliance for Healthy Cities, http://www.alliance-healthycities.com/htmls/members/index_members.html.

Vo Minh Phuc, Vo Trong Cang (2021). Review on Urban Transportation System and Its Adverse Impacts to Public Health of the Cities. *International Journal of Transportation Engineering and Technology. Special Issue: Transportation Engineering Technology and Education. Vol. 7, No. 1, 2021, pp. 24-28*. doi: 10.11648/j.ijtet.20210701.14

Ho Quoc Bang, et al (2017). *"Thực Trạng Ô Nhiễm Không Khí Tp.HCM Và Dự Đoán Xu Hướng Trong Thời Gian Tiếp Theo"* – The Situation of Air Pollution in Ho Chi Minh City and Prediction of Future Trends. Ho Chi Minh City National University and Institute of Environment and Resources (IER).

Le TG., Ngo L., Mehta S., et al. (2012). Effects of Short-term Exposure to Air Pollution on Hospital Admissions of Young Children for Acute Lower Respiratory Infections in Ho Chi Minh City, Vietnam. Health Effects Institute, Boston, MA.

Zhang K, Batterman S (2013). Air pollution and health risks due to vehicle traffic. *Sci Total Environment 2013*; 450-451:307-316. doi: 10.1016/j.scitotenv.2013.01.074.

Benson P (1989). CALINE4 — A Dispersion Model for Prediction Air Pollutant Concentrations Near Roadways. *FHWA-CA-TL-84-15*. Sacramento, CA: California Department of Transportation.

Tien Luc (2020). *"Cải thiện giao thông TP.HCM: Xóa dần "điểm đen" giao thông"* - Improving traffic condition in Ho Chi Minh City: Gradually removing traffic 'black spots'. Retrieved date July 6th, 2021, Ho Chi Minh City Media Center (HMC), https://ttbc-hcm.gov.vn/cai-thien-giao-thong-tphcm-xoa-dan-diem-den-giao-thong-13587.html.

JICA research group (2008). Study on Master Planning of Traffic Safety for Road Traffic in Vietnam – Midterm report. National Traffic Safety Committee (NTSC) and Japan International Cooperation Agency (JICA).

UNESCAP Transport Division (2014). Road Safety in Asia Pacific Region. Regional Expert Group Meeting on Sustainable and Inclusive Transport Development. Ahmedabad, India.

Tu Giang (2020). *"Chiến lược quốc gia bảo đảm trật tự an toàn giao thông đường bộ"* - National strategy to ensure road traffic order and safety. Retrieved date 19/1/2021, *Communist Party of Vietnam Online Newspaper*, https://dangcongsan.vn/thoi-su/chien-luoc-quoc-gia-bao-dam-trat-tu-an-toan-giao-thong-duong-bo-570029.html.

ADB (2015). Viet Nam Urban Environment Program - Urban Sanitation Issues in Viet Nam.

Long Hai Yen and Nguyen Thi Ngoc Anh (2021). Urban Flooding in HCMC. Retrieved date 30/3/2021, ADPC – Asian Disaster Preparedness

Center, https://www.adpc.net/igo/contents/iPrepare/iprepare-news_id.asp?ipid=235#:~:text=The%20first%20reason%20of%20urban,than%20countries%20from%20other%20areas.&text=Additionally%2C%20ICEM%20predicts%20that%20sea,60%25%20of%20HCMC%20in%202050.

World Bank (2021). Projects and Operations. Retrieved date 29/3/2021, World Bank, https://projects. worldbank.org/en/projects-operations/projects-list?countrycode_exact=VN&sector_exact=Sanitation^Public%20Administration%20-%20Water%2C%20Sanitation%20and%20Waste%20Management^Water%20Supply.

Ha An (2020). Saigon vulnerability to flooding worsens. Retrieved date 30/3/2021, VnExpress, https://e.vnexpress.net/news/news/saigon-vulnerability-to-flooding-worsens-4181112.html.

Mook Bangalore, et al (2016). Climate Change and Poverty in Vietnam: Exposure to Floods, Climate Change, and Poverty in Vietnam. Climate Policy Team, World Bank.

Ha Mai (2021). *"Hơn 4 tỉ USD cho giao thông TP.HCM*" – More than 4Billions USD for HCMC transport development. Retrieved date July 6th, Thanh nien Magazine, https://thanhnien.vn/tai-chinh-kinh-doanh/hon-4-ti-usd-cho-giao-thong-tphcm-1351855.html.

AHC (2014). Members of the Alliance for Healthy Cities. Alliance for Healthy Cities Directory 2014. Retrieved date Apr 26th, 2021, Alliance For Healthy Cities, http://www.alliance-healthycities.com/htmls/resources/index_resources.html.

A TRUCK SCHEDULING PROBLEM FOR MULTI-CROSSDOCKING SYSTEM WITH METAHEURISTICS

Duong Bao Tram[1] and Phan Nguyen Ky Phuc[1]

[1] *Department of Industrial Engineering and Management, International University – HCMU, Vietnam, baotramduong99@gmail.com*

[1] *PhD, Department of Industrial Engineering and Management, International University – HCMU, Vietnam, pnkphuc@hcmiu.edu.vn*

Abstract. The subject of scheduling in crossdocking system has always been a recurring matter. Crossdocking system is the result of the effort in optimization in cost and competitiveness amidst an influx of competitors. Though well documented, truck scheduling in crossdocking has yet been implemented much, largely due to the failure in tackling large-sized problems with careful consideration to real-life constraints. In this paper, we focus on a factual problem faced by Son Kim Retail Corporate concerning truck scheduling at the multi-door, multi-crossdocking network with inventory constraints and process capability constraints. An exact mathematical model using mixed-integer linear programming (MILP) is developed, with the objective of minimizing the makespan. The model is inspired by the scheduling model Reservation with Slack of Pinedo, M. (2009) and is solved with the assistant of IBM ILOG CPLEX Optimization Studio. A larger-sized, more complex dataset is dealt with using two metaheuristics, namely Simulated Annealing (SA) algorithm and Tabu Search (TS) algorithm, both of which adhere to the complexity in feasibility assessment of the problem. Performance will be benchmarked between CPLEX and two metaheuristics for small-sized dataset. For large-sized problem, design experiments are made to see how hyperparameters affect the stochastic nature of SA and benchmark the performance between the two metaheuristics. In addition, sensitivity analysis is carried out to observe the dynamic between the main factors to draw meaningful conclusion.

Keywords: Truck scheduling; multi-door, multi-crossdocking network; MILP; CPLEX; Simulated Annealing; Tabu Search

1. Introduction

As the global markets on supply chain has seen an influx of competitors during the past few years, it is pertinent that manufactures, retailers and distributors strive to optimize costs to increase their competitiveness. Driven by such demand, the idea of cross docking was hailed. It was defined by Ladier & Alpan (2016) that cross-docking is a process dealing with transhipping inventory, in which goods and products are unloaded from an inbound truck and process through a flow-center to be directly loaded onto an outbound truck

Domestically, the model of cross-docking in Vietnam has gradually gaining its ground. However, in comparison to the traditional warehouses and distribution centres, it remains meagre in number. Its application is still limited and usually restricted to food industry. One of the prominent retailers using such model is CoopMart, with its cross-docking centre only as an extended part of a bigger distribution centre to facilitate the distribution of perishable and fresh products

1.1. Case study

Son Kim Retail Corporate, specializing retailing, import export and distribution, is a fairy new company. The rise in demand for products in convenient stores and dining restaurants propel the application of a multi-cross docking system with the capacity to deal with multiple types of products. In this paper, the study will focus on the cross-docking process of Son Kim Corporate, which involve separate, multiple docks that have the capability to handle different types of products, some of which require special care like frozen food or fresh seafood.

1.2. Problem statement

The problem faced by Son Kim Corporate is how to coordinate the product and inbound/outbound transportation to the cross-dock in a smooth, time-efficient way, especially when they have to deal with such a large product quantity, type and number of trucks. To elaborate in the detail the problem, with a given number of delivering truck and shipping truck, whose in-transit inventory and demand is known, and a crossdocking systems with known numbers of cross docks, configurations and capacity, the problem is to devise a scheduling plan to come up with the sequence of docks the trucks need to visit, when, and how many products to unload or load that serves to minimize the makespan.

The company needs a solution that makes sure that:

- All demands are satisfied
- The process of successfully loading products onto the shipping truck are done in the shortest amount of time.
- The process of loading and unloading cannot violate the storage capacity and inventory constraints.

1.3. Objective

With respect to the case study, the aim of this paper is to devise a mathematical model as well as a suitable approach for solution development to obtain the aim and satisfy all requirements from the company. The model should reflect on the real conditions to a certain extent to acquire a level of applicability, which can serve as a foundation for future development.

As for the academic aspect, though cross-docking problem has been rigorously explored since the 90s, scheduling problems do not take much proportion in the literary vault. For problem regarding multi-cross dock alone, there have only been two papers publicly released. Paired with the fact that the application of multi cross-dock is quite scarce, there is nearly zero research about the subject conducted in Vietnam, to the best of my knowledge. Therefore, with my model that consider the real-world state and constraint of a Vietnamese business, it can contribute to the existing literature

1.4. Scope

The scope of the problem will fall within the spectrum of crossdocking operational planning through scheduling. However, it will only concern with the exterior operations involving coordinating the trucks, the unloading and loading. The interior operations like material handling, debulking and redistributing will not be concerned. Because it is a short-term planning, the planning horizon will be daily planning for the operation lasting from 22:00 to 5:00, which complies with the transportation legislation for large trucks.

In addition, the methodology subjected to this study include mix-integer linear programming (MILP) and metaheuristics, specifically TS and SA.

An outline of the remainder of this paper is as follows. In Section 2, a literature review of cross-docking problems is presented. Section 3 describes the mathematical formulation of model. Section 4 includes the discussion of TS and SA's implementation, as well as the pseudocode. Section 5 presents the experimental design and Finally, Section 6 draws the conclusions of the research and suggests possible further work.

2. Literature review

One of the earliest works addressing short-term scheduling in cross-dock belongs to Yu (2002), whose work is renowned for considering 32 models. The general aim was to generate a sequence of receiving inbound and outbound truck at each door to minimize the makespan. Later on, Yu and Pius (2008) applied and concluded that the TS metaheuristics could effectively solve a cross-docking problem.

The basic model studied by Baptiste & Maknoon (2007) focused on generating sequences of trucks entering the door, and solutions are developed based on different assumptions to the problem. It was concluded that heuristic algorithm performed better or as well as dynamic programming with or without prior assumption of known truck sequencing. Boysen, Fliedner & Scholl (2008) tackled a simplified cross-dock model to gain the underlying complexity of truck sequencing problem, which was split into subproblems of inbound and outbound sequencing separately.

For multi-door cross-dock, a few studies have taken the approach similar to that of flowshop with parallel machines. Chen & Song (2009) are one of the first people to take such route. Similarly, Nogueira, Coutinho, Ribeiro & Ravetti (2020) also applied the idea to their cases with the addition of time-indexed variables. The problem was approached by using constructive polynomial-time algorithm and more traditional scheduling algorithm like Johnson's rule-based LPT algorithm. Wang & Alidaee (2019) suggested a model for truck sequencing and scheduling in a multi-floor cross-docking center using MINLP.

The literatures regarding truck scheduling are quite well documented over the year. McWilliams et.al (2005) was the first who contributed his work to this topic's literature. The problem, with the objective of minimizing the makespan, was solved using genetic algorithm in combination with simulation model. Another study with regards to industrial requirements is that of Boysen (2009), having been mentioned above.

Golshahi-Roudbaneh, Hajiaghaei-Keshteli, and Paydar (2019) investigated the truck scheduling problem with constraint of time window and deadline for truck departure. Though the paper successfully touched on real-world constraint, the model was simplified to only cross dock with single inbound/outbound door, which may not be applicable to a real situation. The methodology was developed using a hybrid metaheuristic between SA and Keshtel algorithm. Ozden and Saricicek (2019) also tackled the truck scheduling problem with time window constraints but expanded the problem to multi-door cross docking system. The objective was minimizing tardiness of outbound truck and proposed TS and SA for generating the solution.

Conducting truck scheduling at multi-door crossdocking system is a big topic, consisting of works from Wisittipanich & Hengmeechai (2017), Cota et.al (2019), Fonseca et.al (2019), Nogueira et.al (2020). The work of Cota et.al was constructed predicated on the work of Chen & Song (2009) on the two-stage hybrid crossdocking scheduling. The new work appends that of Chen & Song (2009) as the authors used time-indexed model as opposed to the original completion time and precedence model. The study went on to develop the solution using Compressed Differential Heuristic and compared the result coined from both models.

The problem continued to be expanded to multi cross-docking system, or cross-dock network. Madani-Isfahani et al (2014) presented their work on multi-cross dock which intimately adhered to the previously proposed notations by Yu (2002), therefore shared similarity to that of Wisittipanich & Hengmeechai (2017). The work approached the problem of truck scheduling by using sequencing variable. The limitation, however, is the failure to regard the capacity and the increased complexity from the approach. The problem was solved using firefly and SA metaheuristics. The most recent work was that of Correa Issi, Linfati & Escobar (2020), in which the problem of truck scheduling to find the minimum makespan was done for a multi-serviced/purposed crossdocking network. The problem was solely approached by devising a MILP model.

Regarding the similarity of the environment, settings and configuration of the truck scheduling problem and crossdocking network, my work bears some resemblance to the work of Correa Issi, Linfati & Escobar (2020), but expand to multi-door, multi-crossdocking network and Chen et al (2006) in terms of algorithm.

3. Mathematical model

Several assumptions are made to successfully represent the study in the mathematical form.

- The earliest available time of each truck is predetermined by the suppliers
- The inbound doors and outbound doors are separate, meaning each set has single purpose
- The processing time (unloading and loading) depends on the number of products.
- No preemption is allowed
- Shipping truck can only starts loading products when the storage has enough products to satisfy the required loaded amount.
- At all times, each door can only process 1 truck.
- The number of loaded products has to be equal or larger than the demand.
- At the end of the time window, no inventory is allowed.
- The processing time within the cross dock is assumed to be negligible.
- Time starts at 1

Parameters:

R: Number of receiving trucks, r= 1...R
S: Number of shipping trucks, s= 1...S
N: Number of product types, p=1...P
D: Number of cross docks, d=1...D
t: Discrete time period, t=1...T_max
$\boldsymbol{BI_{rp}}$: Number of unit product p on the receiving truck r
$\boldsymbol{DM_{sp}}$: Number of unit product p needed by the shipping truck s
$\boldsymbol{SR_{r}}$: Soonest time receiving truck r enter any dock d
$\boldsymbol{SS_{s}}$: Soonest time shipping truck s enter any dock d
$\boldsymbol{Cap_{d}}$: Capacity of each dock
$\boldsymbol{Door_{d}}$**:** Number of inbound/outbound doors at each cross dock
$\boldsymbol{\alpha}$: Time unloading/unit item
$\boldsymbol{\beta_{dp}}$: Binary product handling capability
$\boldsymbol{TS}$: Transition time of the truck between the docks
$\boldsymbol{BigM}$: Big M

Variables:

$\boldsymbol{C_{max}}$: Makespan

$\boldsymbol{ER_{rd}}$: Time receiving truck r enter cross dock d and unload
$\boldsymbol{DR_{rd}}$: Time receiving truck r finishes unloading and leave dock d
$\boldsymbol{PR_{rdp}}$: Number of unit item p receiving truck r unloaded at dock d
$\boldsymbol{Z1_{rd}}$ =1 if receiving truck r enter dock d, 0 otherwise
$\boldsymbol{Y1_{rdk}}$ =1 if item p is unloaded by receiving truck r at dock d
$\boldsymbol{UR_{rdt}}$ =1 if $t \geq ER_{rd}$, 0 otherwise
$\boldsymbol{VR_{rdt}}$ =1 if $t \leq DR_{rd}$, 0 otherwise
$\boldsymbol{RR_{rdt}}$ =1 if $ER_{rd} \leq t \leq DR_{rd}$, 0 otherwise, $RR_{rdt} = UR_{rdt} \cap VR_{rdt} \cap Z1_{rdt}$
$\boldsymbol{En1_{rdh}}$ =1 if receiving truck enters dock d and dock h, $d, h \in D$, 0 otherwise
$\boldsymbol{X1_{rdh}}$ =1 if receiving truck enters dock d before dock h, $d, h \in D$, 0 otherwise
$\boldsymbol{ES_{sd}}$: Time shipping truck s enter cross dock d
$\boldsymbol{DS_{sd}}$: Time shipping truck s finishes loading and leave dock d
$\boldsymbol{PS_{sdp}}$: Number of unit item p shipping truck s loaded at dock d
$\boldsymbol{Z2_{sdp}}$ =1 if shipping truck r enter dock d, 0 otherwise, 0 otherwise
$\boldsymbol{Y2_{sdp}}$ =1 if item p is loaded onto the shipping truck s at dock d, 0 otherwise
$\boldsymbol{US_{sdt}}$ =1 if $t \geq ES_{sd}$, 0 otherwise
$\boldsymbol{VS_{sdt}}$ =1 if $t \leq DS_{sd}$, 0 otherwise
$\boldsymbol{RS_{sdt}}$ =1 if $ES_{sd} \leq t \leq DS_{sd}$, 0 otherwise, $RS_{sdt} = US_{sdt} \cap VS_{sdt} \cap Z2_{sdt}$
$\boldsymbol{En2_{sdh}}$ =1 if shipping truck enters dock d and dock h, $d, h \in D$, 0 otherwise
$\boldsymbol{X2_{sdh}}$ =1 if shipping truck enters dock d before dock h, $d, h \in D$, 0 otherwise
$\boldsymbol{Load^{t}_{sdp}}$: Amount of product p being loaded onto shipping truck s at dock d
$\boldsymbol{Unload^{t}_{rdp}}$: Amount of product p being unloaded by receiving truck r at dock d
$\boldsymbol{I_{dpt}}$: Inventory of product p at dock d at time t

The mathematical model is developed as followed

Minimize $\boldsymbol{C_{max}}$

s.t.

$$C_{max} \geq DS_{sd} \quad \forall s, d \qquad 1.$$

$$\begin{cases} Y1_{rdp} \leq \beta_{dp}, Y1_{rdp} \leq Z1_{rd} \\ Y1_{rdp} \geq \beta_{dp} + Z1_{rd} - 1 \quad \forall r, d, p \\ PR_{rdp} \leq BigM * Y1_{rdp} \end{cases} \qquad 2.$$

$$SR_r \leq ER_{rd} + BigM(1 - Z1_{rd}) \; \forall r, d \qquad 3.$$

$$DR_{rd} \geq ER_{rd} + \alpha \sum_{p=1}^{P} PR_{rdp} - 1 - BigM(1 - Z1_{rd}) \;\; \forall r, d \qquad 4.$$

$$DR_{rd} \leq ER_{rd} + \alpha \sum_{k=1}^{K} PR_{rkd} - 1 + BigM(1 - Z1_{rd}) \;\; \forall r, d \qquad 5.$$

$$\begin{cases} En1_{rdh} \leq Z1_{rd} \\ En1_{rdh} \leq Z1_{rh} \qquad \forall r \in R, d, h \in D \\ En1_{rdh} \geq Z1_{rd} + Z1_{rh} - 1 \end{cases} \qquad 6.$$

$$ER_{rh} \geq DR_{rd} + TS - BigM(1 - En1_{rdh}) \;\; \forall r \in R, d, h \in D \qquad 7.$$
$$X1_{rdh} + X1_{rhd} = En1_{rdh} \;\; \forall r \in R, d, h \in D$$

$$\sum_{d=1}^{D} PR_{rdp} = BI_{rp} \; \forall r, p \qquad 8.$$

$$\begin{cases} t \geq ER_{rd} - BigM(1 - UR_{rdt}) \\ t \leq ER_{rd} - 1 + BigM * UR_{rdt} \\ t \leq DR_{rd} + BigM(1 - VR_{rdt}) \\ t \geq DR_{rd} + 1 - BigM * VR_{rdt} \\ RR_{rdt} \leq UR_{rdt}, RR_{rdt} \leq VR_{rdt}, RR \leq Z1_{rd} \\ RR_{rdt} \geq UR_{rdt} + VR_{rdt} + Z1_{rd} - 2 \end{cases} \quad \forall r, b, t \qquad 9.$$

$$\sum_{d=1}^{D} RR_{rdt} \leq 1 \quad \forall r, t \qquad 10.$$

$$\textstyle\sum_{r=1}^{R} RR_{rdt} \leq Door_d \quad \forall d, t \qquad 11.$$

$$\begin{cases} Y2_{sdp} \leq \beta_{dp}, Y2_{sdp} \leq Z2_{sd} \\ Y2_{sdp} \geq \beta_{dp} + Z2_{sd} - 1 \quad \forall s, d, p \\ PS_{sdp} \leq BigM * Y2_{sdp} \end{cases} \qquad 12.$$

$$SS_s \leq ES_{rd} + BigM(1 - Z2_{rd}) \; \forall s, d \qquad 13.$$

$$DS_{sd} \geq ES_{sd} + \alpha \sum_{p=1}^{P} PS_{sdp} - BigM(1 - Z2_{sd}) \;\; \forall s, d \qquad 14.$$

$$DS_{sd} \leq ES_{sd} + \alpha \sum_{p=1}^{P} PS_{sdp} + BigM(1 - Z2_{sd}) \quad \forall s, d \qquad 15.$$

$$\begin{cases} En2_{sdh} \leq Z2_{sd} \\ En2_{sdh} \leq Z2_{sh} \qquad \forall s \in S, d, h \in D \\ En2_{sdh} \geq Z2_{sd} + Z2_{sh} - 1 \end{cases} \qquad 16.$$

$$ES_{sh} \geq DS_{sd} + TS - BigM(1 - X2_{sdh}) \;\; \forall s \in S, d, h \in D \qquad 17.$$
$$X2_{sdh} + X2_{shd} = En2_{sdh} \;\; \forall s \in S, d, h \in D$$

$$\sum_{d=1}^{D} PS_{sdp} = DM_{sp} \quad \forall s, p \qquad 18.$$

$$\begin{cases} t \geq ES_{sd} - BigM(1 - US_{sdt}) \\ t \leq ES_{sd} - 1 + BigM * US_{sdt} \\ t \leq DS_{sd} + TS - 1 + BigM(1 - VS_{sdt}) \\ t \geq DS_{sd} + TS + BigM * VS_{sdt} \\ RS_{sdt} \leq US_{sdt}, \quad RS_{sdt} \leq VR_{rdt}, RS_{sdt} \leq Z2_{sd} \\ RS_{sdt} \geq US_{sdt} + VS_{sdt} + Z2_{sd} - 2 \end{cases} \quad \forall s, b, \qquad 19.$$

$$\sum_{d=1}^{D} RS_{sdt} \leq 1 \quad \forall s, t \qquad 20.$$

$$\sum_{s=1}^{S} RS_{sdt} \leq Door_b \quad \forall d, t \qquad 21.$$

$$\begin{cases} \boldsymbol{Unload^t_{rdp} \le BigM*(1-VR_{rdt})} \\ \boldsymbol{Unload^t_{rdp} \le PR_{rdp} + BigM*(VR_{rdt})} \quad \forall r,d,p,t \\ \boldsymbol{Unload^t_{rdp} \ge PR_{rdp} - BigM*(VR_{rdt})} \end{cases} \quad 22.$$

$$\begin{cases} \boldsymbol{Load^t_{sdp} \le BigM*(US_{sdt})} \\ \boldsymbol{Load^t_{sdp} \le PS_{sdp} + BigM*(1-US_{sdt})} \quad \forall s,d,p,t \\ \boldsymbol{Load^t_{sdp} \ge PS_{sdp} - BigM*(1-US_{sdt})} \end{cases} \quad 23.$$

$$\boldsymbol{I_{bkt} = \sum_{r=1}^{R} Unload^t_{rdp} - \sum_{s=1}^{S} Load^t_{rdp}} \quad \forall d,p,t \quad 24.$$

$$\boldsymbol{\sum_{p=1}^{P} I_{dpt} \le Cap_d} \ \forall t,d \quad 25.$$

$$\boldsymbol{\sum_{k=1}^{P} I_{dpt} = 0} \ \forall d, t = T_{max} \quad \forall s,d \quad 26.$$

$$\boldsymbol{C_{max}, ER_{rd}, DR_{rd}, PR_{rdp}, Z1_{rd}, X1_{rdp},, ES_{sdp}, DS_{sd}, PS_{sdp}, Z2_{sd}, X2_{sdp}, I_{dpt} \ge 0} \quad 27.$$

The objective function aims to minimize the makespan. Constraint (1) indicates that makespan has to be equal or greater than all departure time of the shipping trucks. Constraints (2) and (12) depict the dependence of the unloading/loading amount on the decision to visit the dock and the dock's capability. Constraints (3), (4), (5) and (13), (14), (15) define the entering time and departure time based on the earliest available time and the processing time. Constraints (6), (7) and (16), (17) are dock sequential constraints of each truck. Constraint (8) and (18) indicates that all the unloading amounts must be equal to the initial in-transit inventory and all demands must be met. Constraints (9), (10), (11) and (19), (20), (21) are occupation constraint to signify when the docks are occupied and that no truck can enter during occupation time. Constraints (24), (25) and (26) are inventory level constraints, indicating that at no time can the capacity be exceeded.

4. Metaheuristic

4.1. Initial solution

The initial solution is found using the greedy approach. Below is the pseudo code describing the approach.

R: number of receiving truck
S: number of shipping truck
D: number of docks
P: number of products
Refin =[]
Shipfin =[]
Calculate gap between the unloading target amount and available unloading capacity
*Select **r** truck greedily with smallest positive gap and earliest available time*
*Assign **r** truck to docks sequentially*
*Refin.push **(r)***
While shipfin ≤ S
    Find CanShip list that can be satisfied by current inventory
    If ConShip list = []
        Calculate gap between the unloading target amount and available unloading capacity
        Select r truck greedily with smallest positive gap and earliest available time
        Assign r truck to docks with earliest possible assigning time as priority %time that allow unloading with exceeding capacity
    Else
        Calculate gap between the loading target amount and available inventory
        *Select **s** truck greedily with smallest positive gap and earliest available time*
        *Assign **s** truck to docks with earliest possible assigning time as priority %time that allow loading without making inventory below 0*
    End if
End while
Cmax = max (departure time of all s trucks)
Return cmax

4.2. Neighbourhood search

The reason why TS and SA are chosen is due to its simplicity in the process of creating new solution in the process of exploring and exploiting, and the foundation of such process is the neighbourhood search method. In this paper, two swapping methods are looked into.

1	2	3	4	5	6	7

2	1	3	4	5	6	7

Swap 1

1	2	3	4	5	6	7

4	2	3	1	5	6	7

Swap 2

Figure 1: Illustration of 2 swapping methods

The object of swapping is the sequence of receiving trucks, the sequence of shipping trucks and the sequence of docks for the first truck. These are also the input into the evaluation function to calculate the corresponding makespan.

4.3. Evaluation function

The evaluation function follows the same flow as greedy approach. The difference lies within the process of picking trucks to assign. The order off assigning trucks would follow the input sequences. For receiving truck, after a truck is picked, the gap between the unloading target amount and available unloading capacity is calculated. An infeasible input will yield a negative gap and resequencing will be carried out by delaying the assignment of that tuck to the last and consider the next truck in the sequence. Same for the shipping truck, the assigned truck according to the sequence will be compared against the *CanShip list*. If the truck is not in the list, the same process of resequencing will be conducted.

4.4. Tabu Search

Below is the pseudocode of the TS:

```
Create initial solution S
Global_best=S
Best_candidate=S
Tabulist.push(S)
While (loop ≠ max_loop)
    sMove=Create_neighbormove(best_conadidate)
  initialize candidate_list
    if create_neighbor(sMove) not in tabulist
             candidate_list= create_neighbor(sMove)
    end
  if candidate_list = [] or global_best is better than
  target
        break loop
  end
    let best_candidate = candidate_list[1]
    for (S' in candidate_list)
          if find_cmax(S') <
find_cmax(best_candidate)
                best_candidate=S'
          end
    end
    tabulist.push(best_candidate)
    if find_cmax(best_candidate) <
find_cmax(global_best)
             global_best = best_candidate
    end
end.
```

4.5. SA (Sigmoid function)

Below is the code of SA using Sigmoid function

```
Create initial solution S
Current = S
Next_move = empty
run =1
While T > mint
    While run <= numrun
        sMove=Create_neighbormove(best_conadidat
        e)
        select a movement randomly
        next_move = selected movement
        delta = current – mext_move
        P = 1/(1+e^(-del /T)) %Sigmoid function
        If chances to move < P
            Current = next_move
            Run =run+1
        Else
            Eliminated the move
        end
    end
    Decrease(T)
end.
```

4.6. SA (Metropolis function)

The difference between the approach using Metropolis function and Sigmoid function is the way the stochastic movement between the current solution and a new solution is carried out. For Sigmoid function, every movement adheres to the probability, which depends on the delta and temperature *T*. For Metropolis function, the movement is carried out like below:

```
    delta = eval(next move)-eval(current)
    If delta < 0
        Current = next_move    %chances to move is
        100%
        Run = run+1
    Else
        P = e^(-delta/T)
        If chance to move < P
            Current = next_move
            Run=run+1
        Else
            Eliminate the move
        end
    end
```

5. Result analysis

A total of 10 data sets are considered in this paper, ranging from small-scaled to large-scaled problem.

Table 1: Data sets' information

Problem	Retruck	Shiptruck	Product	Total quantity	Nature
1	3	4	4	140	not random
2	3	4	5	64	random
3	3	4	8	180	random
4	3	4	9	188	random
5	3	4	10	444	not random
6	5	4	6	1030	random
7	6	4	8	491	random
8	5	7	10	430	not random
9	6	7	10	2020	random
10	6	11	15	1252	not random

The following parts analyse the performance of each approach based on the results of the data sets.

5.1. CPLEX's performance

Table 2: CPLEX's performance

Problem	Cmax	Run time
1	112	20
2	36	37
3	110	57
4	110	81
5	277	217
6	476	9478
7	221	9193
8	208	10812
9	Unsolved	
10	Unsolved	

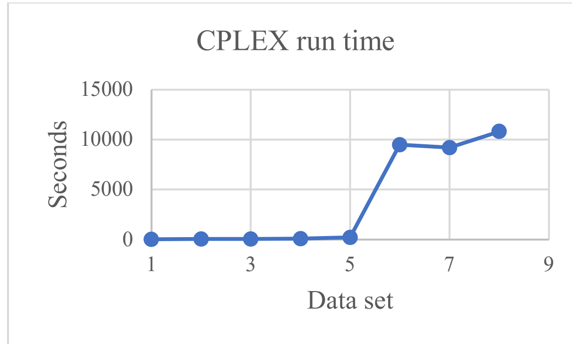


Figure 2: Run time of CPLEX

For small-scaled problems with small product quantity and number of trucks, CPLEX works quite well in terms of run time, which only takes less than 2 minutes for dataset 1, 2 and 3 and 4.

The run time exponentially increases at the data set 6, whose changes include the increase in all fundamental factors of the problem. This is explained by the increase in the number of trucks and product quantity, which makes the number of variables exponentially grow in number and therefore makes the problem more cumbersome and complex.

5.2. Tabu Search's performance

Because the way TS find the best solution is to search across the neighbourhood (NBH), Neighbourhood Search method has a large influence on the performance of the algorithm. Table 5.3 displays the results from two swapping method.

Table 3: TS's performance

Dataset	NBH size	Cmax	Run time	NBH size	Cmax	Run time
1	20	112	25.61	42	112	28.56
2	36	36	24.21	112	36	92.6
3	36	114	491.2	112	114	138.35
4	32	115	453.48	84	115	122.72
5	12	313	16.57	42	299	325.43
6	60	524	940.45	308	519	1934
7	28	282	27.1	413	242	231.1
8	35	214	881.86	246	209	1542.9
9	119	1076	2113.3	1364	1026	3197.9
10	66	629	1710.8	1568	627	3099.9

Firstly, concerning the NBH size, due to having more degree of freedom in the swapping method, the number of NBH sizes for method 2 exponentially proliferate compared to those of method 1. Method 1 also creates 2 outliers in data set 5 and 7 which are caused by early convergence. Method 2 also consistently yields better or as good results as method 1 does for every data set.

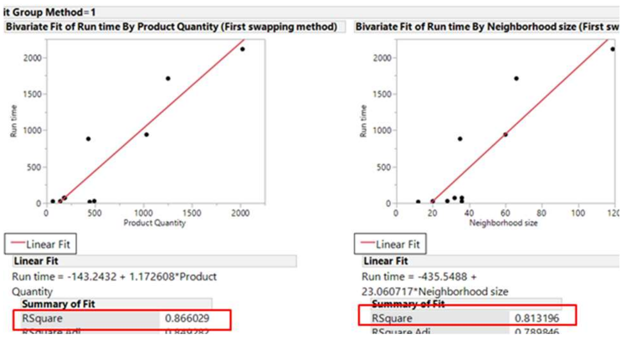


Figure 3: Correlation between run time, NBH size and product quantity (Method 1)

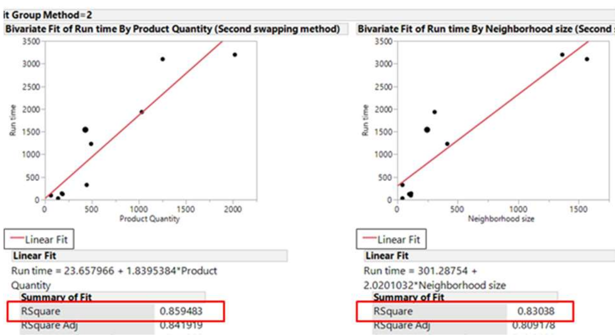


Figure 4: Correlation between run time, NBH size and product quantity (Method 2)

The figures using JMP analysis also gives more information on the correlations between run times and factors like product quantity and NBH size. For both methods, these two factors strongly affect the run time, with the correlation coefficient between run time and product quantity 0.93 for method 1 and 0.92 for method 2, while that with NBH size being 0.9 for method 1 and 0.91 for method 2.

5.3. Simulated Annealing's performance

As mentioned in section 3, this paper discussed two approaches to the implementation of SA, including Sigmoid function and Metropolis function. First, experimental designs are run to choose the hyperparameters' value

5.3.1. Hyperparameters' experimental design

The experimental design for both approaches is carried out in the similar manner. The experimental runs are conducted on the first 5 data sets, with the aim of being able to benchmark the result with CPLEX. The following table shows different values of the hyperparameters for the experimental runs.

Table 4: Hyperparameters' value for the experimental design

Factor	Explanation	Value
numloop	the number of run loop	30
		20
numrun	the number of sub-iterations	20
		10
T_decrease	temperature changing constant	0.8
		0.9
		0.99

The assessment indices for the performance of each combination are the run time, and delta – the gap between the result of SA and CPLEX. Below are the variable charts for both SA sigmoid and SA Metropolis.

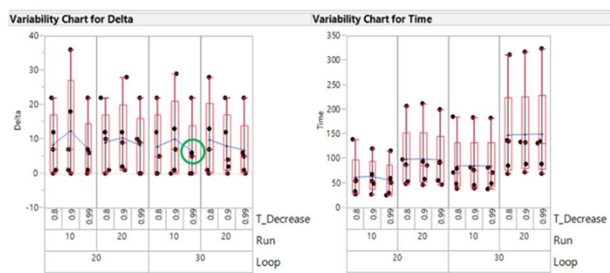


Figure 5: Variability chart of SA Sigmoid's Delta and run time

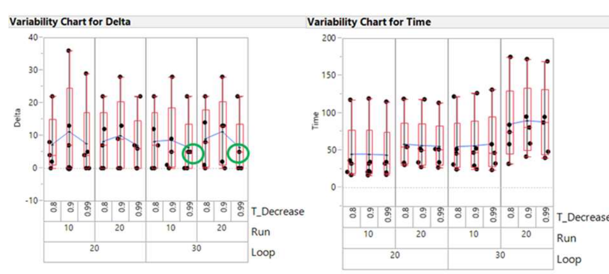


Figure 6: Variability chart of SA Metropolis's Delta and run time

For both Metropolis function and Sigmoid function, the combination producing the minimum delta as well as having the acceptable run time are *T_decrease* =

0.99, *numloop* =30, *numrun* = 10, among which *T_decrease* was the factor having the prominent impact on the delta, while run time was mainly affected by *numloop* and *numrun.*

5.3.2. Performance of SA

As SA depends on not just the NBH Search method, but also the stochastic nature in moving between points in the NBH, the dynamic of the algorithm and the factors is postulated to be different than that with TS. The table below shows the performance of two approaches using two different swapping method.

Table 5: SA Sigmoid's performance

Dataset	NBH size	Cmax	Run time	NBH size	Cmax	Run time
1	20	112	49.39	36	112	52.01
2	36	36	38.06	112	36	40.47
3	36	116	71.47	112	120	61.61
4	32	115	81.67	98	115	80.83
5	12	299	182.68	120	313	172.42
6	60	551	298.92	308	524	300.60
7	28	251	170.90	413	258	160.80
8	35	212	335.67	246	223	290.93
9	119	1088	928.21	1364	1102	912.55
10	121	630	853.39	1568	630	843.07

Table 6: SA Metropolis' performance

Dataset	NBH size	Cmax	Run time	NBH size	Cmax	Run time
1	20	112	32.30	42	112	33.04
2	36	36	23.64	112	36	23.56
3	36	115	46.56	112	117	38.76
4	32	115	58.08	84	115	53.32
5	12	299	131.09	42	299	104.27
6	60	551	194.82	308	540	184.80
7	28	261	104.87	413	242	112.33
8	35	212	198.05	246	213	183.23
9	119	1026	553.30	1364	1026	559.82
10	121	629	636.00	1568	630	562.19

It can be seen that, for both Metropolis and Sigmoid, though the run time lengthens as the complexity increases, the drastic change in the NBH sizes caused by different swapping method does not lead to changes in the run time like the case of TS. The results yielded by method 2 also did not make a significant improvement from method 1. Concerning the product quantity, from the JMP analysis, the correlation coefficient between the run time and the quantity is only 0.71 for both approaches, which is a quite loose correlation.

Figure 5.6 shows the comparison of the results between two SA's approaches.

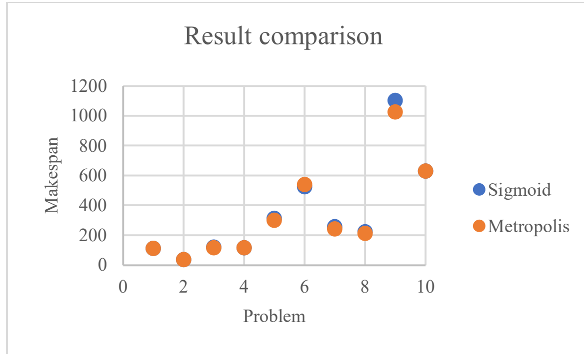


Figure 7: Result comparison between Metropolis and Sigmoid

It can be seen that 9/10 times, the results from Metropolis function are better or on par with those of Sigmoid function. In addition, based on table 5.5 and 5.6, the run time of Metropolis is 1.6 times faster than that of Sigmoid on average. Metropolis also exhibit a higher likelihood of achieving the *globalbest* result with 2.5/30 loops having the best results compared to the 1.7/30 loops of Sigmoid. Therefore, it can be concluded that Metropolis function is a better approach to implementing SA.

5.4. Comparison between CPLEX and the metaheuristics

Table 7: Performance gap between CPLEX and the metaheuristics

Problem	Performance gap (Benchmark with Cplex)		
	TS	SA (Sigmoid)	SA (Metropolis)
1	0.00%	0.00%	0.00%
2	0.00%	0.00%	0.00%
3	3.64%	5.45%	4.55%
4	4.55%	4.55%	4.55%
5	7.94%	7.94%	7.94%
6	9.03%	10.08%	13.45%
7	9.50%	13.57%	9.50%
8	0.48%	1.92%	1.92%

On an overall viewpoint, TS's results yield a smaller gap to CPLEX than SA does. 62.5% of the TS's results exhibit the gap below 5%, while the statistic of SA is 50%. None of TS's gaps are over 10%, while Metropolis function produces 1 case and Sigmoid 2 cases. Therefore, a clear compromise between run time and good results are observed between the two approaches.

6. Conclusion and suggestions for future work.

In conclusion, to solve the problem of truck scheduling in crossdocking network, 3 approaches are taken. The first is using MILP in conjunction with CPLEX to solve for the exact solution. However, because of its restriction to small-sized problems, TS and SA are implemented to search for the makespan of large-sized problems. The two metaheuristics exhibit the tradeoff between producing a consistent and good result and having short run time.

The result also proves the credibility and feasibility of the model as well as the algorithm. Regarding the all-encompassing and real-life adherent nature of the proposed model, not only does it make a solid contribution to the topic's literature but also serve as a foundation for further development of the program into software.

Further study on this topic can be expanded to include the interior operations of the crossdocking network. Another direction is to expand the problem downstream by combining the truck scheduling problem with the vehicle routing problem to the customers.

Reference

Alharkan, I.M. (2005), Algorithms for sequencing and scheduling, p.5-11.

Baptiste, P., & Maknoon, M. Y. (2007), Cross-docking: Scheduling of incoming and outgoing semi-trailer. In International conference on production research –ICPR'07. Valparaiso, Chile.

Berghman, L., Briand, C., Leus, R., & Lopez, P. (2015). The Truck Scheduling Problem at Crossdocking Terminals - Exclusive versus Mixed Mode. Proceedings Of the International Conference on Operations Research and Enterprise Systems. https://doi.org/10.5220/0005205102470253

Boysen, N. (2010). Truck scheduling at zero-inventory cross docking terminals. Computers & Operations Research, 37(1), 32-41. https://doi.org/10.1016/j.cor.2009.03.010

Boysen, N., & Fliedner, M. (2010). Cross dock scheduling: Classification, literature review and research agenda. Omega, 38(6), 413-422. https://doi.org/10.1016/j.omega.2009.10.008

Buakum, D. (2019). A Literature Review and Further Research Direction in Cross-docking.

Boysen, N., Fliedner, M., & Scholl, A. (2008). Scheduling inbound and outbound trucks at cross docking terminals. OR Spectrum, 32(1), 135–161.

Chen, P., Guo, Y., Lim, A., & Rodrigues, B. (2006). Multiple crossdocks with inventory and time windows. Computers & Operations Research, 33(1), 43-63. https://doi.org/10.1016/j.cor.2004.06.002

Chen, F., & Song, K. (2009). Minimizing makespan in two-stage hybrid cross docking scheduling problem. Computers and Operations Research, 36, 2066–2073.

Correa Issi, G., Linfati, R., & Escobar, J. (2020). Mathematical Optimization Model for Truck Scheduling in a Distribution Center with a Mixed Service-Mode Dock Area. Journal Of Advanced Transportation, 2020, 1-13. doi: 10.1155/2020/8813372

Cota, P., Gimenez, B., Araújo, D., Nogueira, T., de Souza, M., & Ravetti, M. (2016). Time-indexed formulation and polynomial time heuristic for a multi-dock truck scheduling problem in a cross-docking centre. Computers & Industrial Engineering, 95, 135-143. https://doi.org/10.1016/j.cie.2016.03.001

Fatthi, W. N. A. W. A., Shuib, A., Dom, R. M. & Shahidin, A. M. (2014). Ideology, Decision Planning and Operational Scenarios and Features in Cross Docking Models. American Journal of Applied Sciences, 11(12), 2004-2015. https://doi.org/10.3844/ajassp.2014.2004.2015

Fonseca, G., Nogueira, T., & Ravetti, M. (2019). A hybrid Lagrangian metaheuristic for the cross-docking flow shop scheduling problem. European Journal of Operational Research, 275(1), 139-154. https://doi.org/10.1016/j.ejor.2018.11.033

Glauber, R. J (1963), Time-Dependent Statistics of the Ising Model, J. Math. Phys. 4, 294

Golshahi-Roudbaneh, A., Hajiaghaei-Keshteli, M., & Paydar, M. M. (2019). A hybrid approach in metaheuristics for a cross-dock scheduling considering time windows and deadline for trucks departure. Scientia Iranica.

Heinrici A. (1994) A Comparison between Simulated Annealing and Tabu Search with an Example from the Production Planning. In: Operations Research Proceedings 1993. Operations Research Proceedings (GOR (Gesellschaft für Operations Research e.V.), vol 1993. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-78910-6_162

Jarraya1, B., & Bouri, A. (2012). Metaheuristic Optimization Backgrounds: A Literature Review. International Journal of Contemporary Business Studies. Vol: 3, No: 12. December 2012, ISSN 2156-7506

Kunche, P., & Reddy, K. (2016). Heuristic and Meta-Heuristic Optimization. Metaheuristic Applications to Speech Enhancement, 17-24. https://doi.org/10.1007/978-3-319-31683-3_3

Liao, T., Chang, P., Kuo, R., & Liao, C. (2014). A comparison of five hybrid metaheuristic algorithms for unrelated parallel-machine scheduling and inbound trucks sequencing in multi-door cross docking systems. Applied Soft Computing, 21, 180-193. https://doi.org/10.1016/j.asoc.2014.02.026

Madani-Isfahani, M., Tavakkoli-Moghaddam, R., & Naderi, B. (2014). Multiple cross-docks scheduling using two meta-heuristic algorithms. Computers & Industrial Engineering, 74, 129-138. https://doi.org/10.1016/j.cie.2014.05.009

McWilliams, D., Stanfield, P., & Geiger, C. (2005). The parcel hub scheduling problem: A simulation-based solution approach. Computers & Industrial Engineering, 49(3), 393-412. https://doi.org/10.1016/j.cie.2005.07.002

Nogueira, T., Coutinho, F., Ribeiro, R., & Ravetti, M. (2020). Parallel-machine scheduling methodology for a multi-dock truck sequencing problem in a cross-docking center. Computers & Industrial Engineering, 143, 106391. https://doi.org/10.1016/j.cie.2020.106391

Ozden, G. & Saricicek, I. (2019). Scheduling trucks in a multi-door cross-docking system with time windows. Bulletin of the Polish Academy of Sciences. Technical Sciences 67(2).

Pinedo, M. (2009). Planning and Scheduling in Manufacturing and Services. https://doi.org/10.1007/978-1-4419-0910-7

Wang, H., & Alidaee, B. (2019). The multi-floor cross-dock door assignment problem: Rising challenges for the new trend in logistics industry. Transportation Research Part E: Logistics and Transportation Review, 132, 30-47. https://doi.org/10.1016/j.tre.2019.10.006

Wisittipanich, W., & Hengmeechai, P. (2017). Truck scheduling in multi-door cross docking terminal by modified particle swarm optimization. Computers & Industrial Engineering, 113, 793-802. https://doi.org/10.1016/j.cie.2017.01.004

Yu, W. (2002), Operational strategies for cross docking systems. Dissertation. Iowa State University: Ames, IA, USA.

Yu, W., & Pius, J. E. (2008). Scheduling of inbound trucks in cross docking systems with temporary storage. European Journal of Operational Research, 184(1), 377–396

SOLVING VEHICLE ROUTING PROBLEM OF A FMCG COMPANY IN VIETNAM USING K-NEAREST NEIGHBOR ALGORITHM

Tran Quynh Le[1, *], Thanh-Tuan Dang[1] and Chia-Nan Wang[1]

[1] Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

**Corresponding author: quynhle220292@gmail.com (T.Q.L)*

Abstract. In recent times, the Fast-Moving Consumer Goods (FMCG) industry has encountered numerous challenges to satisfy diverse customer demand. Because of the short shelf life of products, the Vehicle Routing Problem (VRP) figures prominently in FMCG logistics and supply chain systems. Therefore, this study aims to determine the optimal truck routes to minimize the total delivery distance from a distribution center to retailers. A K-nearest neighbor algorithm is adopted to identify the best courses. A network of a dairy company in Vietnam, which has a distribution center, 20 retailers, and three types of trucks, is used to demonstrate the applicability of the model and solution methodology. A sensitivity analysis is conducted to investigate the impact of the truck types under various capacities on the traveling distance and utility. It found that two big trucks, whose standard can carry 36 pallets per one, are suitable for milk products delivering with 96% vehicle capacity utilization.

Keywords: Vehicle routing problem; k-nearest neighbor algorithm; FMCG; distribution center; retail.

1. Introduction

With the development of e-commerce, FMCG has significant growth potential, especially the impact of COVID-19. These goods are purchased frequently, are consumed rapidly, are priced low, and are sold in large quantities. Transportation is an important operation making the products available to the customers. If delivery is delayed, customer satisfaction with the company will be affected. The determination of the transportation route and the type of vehicles in shipping activities is necessary to meet the demand of customers and reduce costs. Basically, there are many actual applications that can deliver goods from distribution to the customer quickly and efficiently. The truck dispatching problem was first considered by Dantzig et al. (1959) in which a set of stations is supplied by a fleet of gasoline delivery trucks from a terminal with optimal routing.

A vehicle routing problem (VRP) is used to determine routes of identical vehicles at one or more depots therefore each customer is visited exactly by drivers, and the total transportation distance is minimized. A vehicle routing network including one or more depots and customers. The demand of customers is satisfied by inventories at the depot. A vehicle is then assigned to ship the commodities to the customers within the loading capacity limit of the vehicle, according to a transportation route. Fig.1 shows an example of a vehicle routing problem with a single depot and customer demands.

This research proposed the delivery optimization route from distributors to the agent retailers of a fast-moving consumer goods company. The distributer's delivery system is presented and an appropriate VRP model in Section 2. Section 3 presents a solution to the problem.

A case study was introduced as an actual implementation in section 4. Section 5 presents the conclusion.

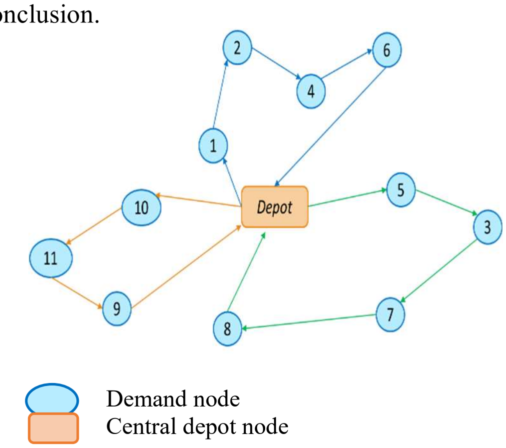


Figure 1. Vehicle routing problem example

2. Literature review

The vehicle routing problem is known as a delivery problem or vehicle scheduling. According to Hanne et al. (2015), determining the schedule and delivery route from distributions to customer nodes is a vital decision in transportation and logistics. The vehicle routing problem is studied and applied in companies, as such as communal home meal delivery service (Bräysy et al. 2009 and Yildiz et al. 2013), livestock collection (Oppen et al. 2008), waste collection (Bautista et al. 2008), Shuttle Service (Pei-Ying et al. 2013), to name a few.

The vehicle routing problem is usually formulated as a linear programming model, and the solution for this problem included exact algorithms, classical heuristics, and metaheuristics. Exact algorithms were proposed by Laporte et al. (1985) and Laporte et al. (1988), the optimal solutions are then found by the branch and bound algorithms. Classical heuristics were developed mostly between 1960 and 1990 (Laporte et al. 2000). Clarke et al. (1964) developed a savings algorithm, which is an iterative procedure that enables the rapid selection of an optimum or near-optimum route. The sweep algorithm is proposed by Gillett et al. (1974) to solve medium- as well as large-scale vehicle routing problems. Balinski et al. (1964), Agarwal et al. (1989), and Renaud et al. (1996) extended the sweep algorithm is to generate petals heuristics to solve the VRP model. Metaheuristics algorithms were designed to solve approximately a wide range of hard optimization problems (Boussaïd et al. 2013). Elshaer et al. (2020) classified VRP articles published between 2009 and 2017 with two categories: single solution-based metaheuristics and population-based metaheuristics.

3. Mathematical model

3.1. Problem description and formulation

The VRP model in this study considers a large network that consists of n nodes. The network has a depot that vehicles start and finish the shipment. A single delivery location has regular orders everyday D_i. Each delivery location must be visited once by a vehicle. And the capacity of the vehicle is not over-utilized C. The objective of this study is to obtain the best delivery route to minimize the transportation distance by using a nearest-neighbor procedure.

Set and parameter of model:

N	Set of nodes (0, 1, …, n). Set of depot O = {0}
K	The number of available vehicles
d_{ij}	Transport distance from node *i* to node j (km)
D_i	The quantity of regular demand at node $i \in N$
C	Vehicle capacity
x_{ij}	binary decision variable that indicates whether the vehicle travels from nodes *i* to *j*

$$Min \sum_{i \epsilon N} \sum_{j \epsilon N} d_{ij} x_{ij} \qquad (1)$$

Subject to

$$\sum_{i \epsilon N} x_{ij} = 1, \qquad \forall j \epsilon N \backslash \{0\} \qquad (2)$$

$$\sum_{j \epsilon N} x_{ij} = 1, \qquad \forall i \epsilon N \backslash \{0\} \qquad (3)$$

$$\sum_{i \epsilon N} x_{i0} \le K \qquad (4)$$

$$\sum_{j \epsilon N} x_{0j} \le K \qquad (5)$$

$$\sum_{i \epsilon O} \sum_{j \epsilon N} x_{ij} \le 1 \qquad (6)$$

$$\sum_{i \epsilon N} \sum_{j \epsilon N} D_i x_{ij} \le C \qquad (7)$$

$$x_{ij} \epsilon \{0, 1\} \quad \forall i, j \in N \qquad (8)$$

The VRP is modeled by using the formulation as (1)-(6). The objective function (1) minimizes the total transportation distance by all vehicles. Constraints (2) and (3) ensure each delivery location is visited by only one vehicle. Constraints (4) and (5) ensure that the number of vehicles leaving the depot is equal to the number entering. Constraints (6) state that each vehicle does not leave the depot more than once. Constraint (7) is the vehicle capacity constraint, that ensures that the total load has to not exceed the vehicle capacity. Constraint (8) is the integer condition for variables.

3.2 The proposed approach: NNP

The nearest neighbor procedure (NNP) is a simple way when compared with other advanced heuristic algorithms. The nearest neighbor procedure builds a tour based only on the cost or distance of traveling from the last-visited node to the closest node in the network. The procedure is outlined as follow:

1. Start with the depot node
2. Find the node closest to the last node added to the tour
3. Go back to step 2 until all nodes have been added

4. Connect the first and the last nodes to form a complete tour

4. Case study

VNM is a large company, it is normal to consume goods faster than expected or agents are in short supply. Every day, the company has a lot of orders coming in, so the distribution of goods between warehouses and the dealers is an important operation that helps to meet the demand of customers at the possible minimum costs. Therefore, the NPP algorithm is applied to improve the operation of the fleet and reduce transportation costs.

VNM company has a total of 21 nodes, including a warehouse and 20 retail stores (Table 1). The distances between any pair of stores are gathered via Google maps. The calculated distance matrix values are shown in Table 2. Every day, commodities are transported from a warehouse to retail stores with different dairy demands. The volume of the dairy demand was collected based on the procurement department. The company has 5 vehicles (2 big trucks and 3 small trucks). Each small truck has a charge of 36 pallets and each big truck has 25 pallets.

Table 1. Customers with location and demand

No	Address	Latitude	Longitude	Demand
Warehouse	32 Dang Van Bi, Thu Duc District, Ho Chi Minh City	10.848	106.760	0
Store 1	31 Dinh Tien Hoang, DaKao, D.1	10.789	106.699	5
Store 2	135 Cong Quynh, Nguyen Cu Trinh, D.1	10.766	106.689	4
Store 3	89 Cach Mang Thang Tam, Ben Thanh, D.1	10.773	106.691	5
Store 4	359 Le Van Sy, Ward 14, D.3	10.789	106.677	4
Store 5	190 Nguyen Dinh Chieu, Ward 6, D.3	10.779	106.690	1
Store 6	1A Truong Son, Ward 15, D.10	10.807	106.664	5
Store 7	494 Ly Thai To, Ward 10, D.10	10.768	106.672	3
Store 8	58 La Xuan Oai, Tang Nhon Phu A, D.9	10.844	106.788	1
Store 9	267A Do Xuan Hop, Phuoc Long B, D.9	10.824	106.769	5
Store 10	10 Tan Trao, Tan Phu, D.7	10.730	106.724	3
Store 11	63 Lam Van Ben, Tan Thuan Tay, D.7	10.748	106.716	4
Store 12	482 Le Quang Dinh, Ward 11, Binh Thanh District	10.814	106.690	3
Store 13	232 Nguyen Xi, D.13, Binh Thanh District	10.818	106.706	4
Store 14	113 An Duong Vuong, An Lạc Ward, Binh Tan District	10.720	106.619	4
Store 15	91 Le Van Quoi, Binh Tri Dong, Binh Tan District	10.775	106.619	3
Store 16	105 Phan Dinh Phung, Ward 17, Phu Nhuan District	10.795	106.684	2
Store 17	979C Kha Vạn Can, Linh Chieu Ward, Thu Duc District	10.851	106.755	3
Store 18	273 Le Van Khuong, Hiep Thanh Ward, D.12	10.873	106.649	1
Store 19	464C Luy Ban Bich, Hoa Thanh, Tan Phu District	10.780	106.636	6
Store 20	384 Pham Hung, Ward 5, D.8	10.737	106.671	3

Table 2. Distance Matrix

	WH	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
WH	-	12.5	15.0	15.0	15.0	14.0	16.9	16.3	5.5	3.1	19.2	18.2	10.1	9.0	26.3	24.5	14.0	1.6	16.0	20.0	22.0
S1	12.5	-	4.6	3.5	3.1	2.4	4.5	4.5	15.1	11.5	9.7	7.8	3.4	5.4	13.1	10.8	1.9	10.4	14.8	9.8	8.3
S2	15.0	4.6	-	1.3	3.8	2.6	4.0	2.2	17.5	13.9	9.0	6.5	7.1	9.1	10.3	8.9	4.3	14.1	17.7	7.9	5.0
S3	15.0	3.5	1.3	-	3.5	1.0	3.8	2.6	15.9	12.3	9.4	6.2	6.6	8.5	10.7	9.3	4.0	13.6	17.4	8.3	5.7
S4	15.0	3.1	3.8	3.5	-	2.2	2.5	3.5	16.9	15.5	11.3	8.7	4.9	8.1	12.0	9.0	2.4	13.0	14.5	7.5	7.2
S5	14.0	2.4	2.6	1.0	2.2	-	3.2	3.1	16.1	13.4	10.1	7.4	5.3	7.3	11.3	9.4	2.6	12.3	16.0	8.4	6.3
S6	16.9	4.5	4.0	3.8	2.5	3.2	-	3.3	18.3	14.7	12.1	9.5	7.8	9.5	10.9	7.4	4.0	14.5	15.2	5.3	6.4
S7	16.3	4.5	2.2	2.6	3.5	3.1	3.3	-	18.1	14.5	9.9	7.4	7.3	9.3	8.8	6.7	4.0	14.3	17.4	5.7	4.0
S8	5.5	15.1	17.5	15.9	16.9	16.1	18.3	18.1	-	4.3	18.9	20.1	13.2	12.1	26.4	24.6	15.8	4.4	18.9	23.0	22.5
S9	3.1	11.5	13.9	12.3	15.5	13.4	14.7	14.5	4.3	-	17.1	16.7	11.2	10.1	22.7	21.0	12.1	4.4	19.0	20.0	18.8
S10	19.2	9.7	7.8	7.7	10.5	9.2	11.1	9.3	19.0	17.1	-	2.8	13.0	14.9	15.6	15.8	11.5	19.8	27.2	14.8	6.6
S11	18.2	7.8	6.5	6.2	8.7	7.4	9.5	7.4	20.1	16.7	2.8	-	10.0	11.0	13.0	14.0	7.9	19.4	21.7	12.8	6.6
S12	10.1	3.4	7.1	6.6	4.9	5.3	7.8	7.3	13.2	11.2	13.0	10.0	-	2.6	16.0	11.0	3.0	9.3	11.6	9.5	10.5
S13	9.0	5.4	9.1	8.5	8.1	7.3	9.5	9.3	12.1	10.1	14.9	11.0	2.6	-	17.0	13.0	6.9	8.2	16.3	14.6	12.5
S14	26.3	13.1	10.3	10.7	12.0	11.3	10.9	8.8	26.4	22.7	15.6	13.0	16.0	17.0	-	8.3	13.0	25.4	21.4	8.7	9.0
S15	24.5	10.8	8.9	9.3	9.0	9.4	7.4	6.7	24.6	21.0	15.8	14.0	11.0	13.0	8.3	-	10.0	19.9	14.7	2.5	9.0
S16	14.0	1.9	4.3	4.0	2.4	2.6	4.0	4.0	15.8	12.1	11.5	7.9	3.0	6.9	13.0	10.0	-	12.1	13.4	7.0	7.7
S17	1.6	10.4	14.1	13.6	13.0	12.3	14.5	14.3	4.4	4.4	19.8	19.4	9.3	8.2	25.4	19.9	12.1	-	14.4	18.0	17.5
S18	16.0	14.8	17.7	17.4	14.5	16.0	15.2	17.4	18.9	19.0	27.2	21.7	11.6	16.3	21.4	14.7	13.4	14.4	-	11.8	20.1
S19	20.0	9.8	7.9	8.3	7.5	8.4	5.3	5.7	23.0	20.0	14.8	12.8	9.5	14.6	8.7	2.5	7.0	18.0	11.8	-	8.2
S20	22.0	8.3	5.0	5.7	7.2	6.3	6.4	4.0	22.5	18.8	6.6	6.6	10.5	12.5	9.0	9.0	7.7	17.5	20.1	8.2	-

Table 3. The best rout generated with NNP

3 Small Trucks **Capacity: 75 pallets**	*1st small truck:* DC → Store 17 → Store 8 → Store 9 → Store 13 → Store 12 → Store 16→ Store 1 → Store 5 → DC **TD = 1.6 + 4.4 + 4.3 + 10.1 + 2.6 + 3 + 1.9 + 2.4 + 14 = 44.3 km**	**44.3 + 53.9 + 73.6 = 171.8 km**
	2nd small truck: DC → Store 2 → Store 3 → Store 7 → Store 6 → Store 4 → Store 20 → DC **TD = 15 + 1.3 + 2.6 + 3.3 +2.5 + 7.2 + 22 = 53.9 km**	
	3rd small truck: DC → Store 18 → Store 19 → Store 15 → Store 14 → Store11→ Store 10 → DC **TD = 16 + 11.8 + 2.5 +8.3 + 13 +2.8 + 19.2 = 73.6 km**	
2 Big trucks **Capacity: 72 pallets**	*1st big truck:* DC → Store 17 → Store 8 → Store 9 → Store 13 → Store 12 → Store 16→ Store 1 → Store 5 → Store 3 → Store 2 → Store 7→ DC **TD = 1.6 + 4.4 + 4.3 + 10.1 + 2.6 + 3 + 1.9 + 2.4 + 1 + 1.3 + 2.2 + 16.3 = 51.1 km**	**51.1 + 89.8 = 140.9 km**
	2nd big truck: DC → Store 4 → Store 6 → Store 19 → Store 15 → Store 14 → Store 20 → Store 10 → Store 11 → Store 18 → DC. **TD = 15 + 2.5 + 5.3 + 2.5 + 8.3 + 9 + 6.6 + 2.8 + 21.8 + 16 = 89.8 km**	
2 Small trucks + **1 Big truck** **Capacity: 86 pallets**	*1st big truck:* DC → Store 17 → Store 8 → Store 9 → Store 13 → Store 12 → Store 16→ Store 1 → Store 5 → Store 3 → Store 2 → Store 7→ DC **TD = 1.6 + 4.4 + 4.3 + 10.1 + 2.6 + 3 + 1.9 + 2.4 + 1 + 1.3 + 2.2 + 16.3 = 51.1 km**	**51.1 + 64.6 + 59.7 = 175.4 km**
	1st small truck: DC → Store 4 → Store 6 → Store 19 → Store 15 → Store 14 → Store 20 → DC **TD = 15 + 2.5 + 5.3 + 2.5 + 8.3 + 9 + 22 = 64.6 km**	
	2nd small truck: DC → Store 18 → Store 11 → Store 10 → DC **TD = 16 +21.7 + 2.8 + 19.2 = 59.7 km**	

In Table 1, the total demand of all retail stores is 69 pallets. Therefore, there are 3 options for selecting the number of vehicles to deliver goods to stores. The first option, company uses 3 small rucks to transport pallets. The second option, they choose the shipment by 2 big trucks. And in the third option, goods are shipped by using 2 small trucks and 1 big truck. The total traveling distances of other options are calculated in Table 3.

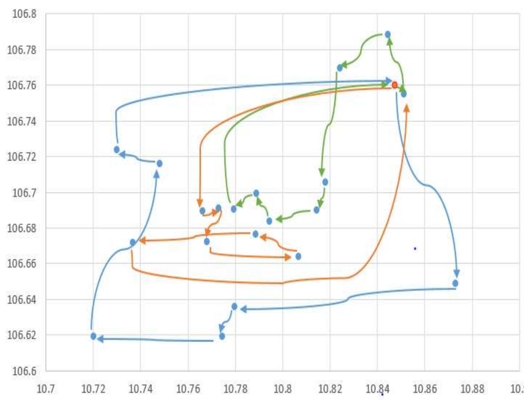


Figure 2. The best route for option 1 (3 small trucks)

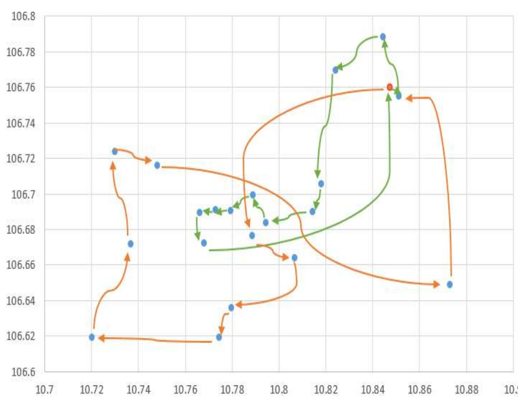


Figure 3. The best route for option 2 (2 big trucks)

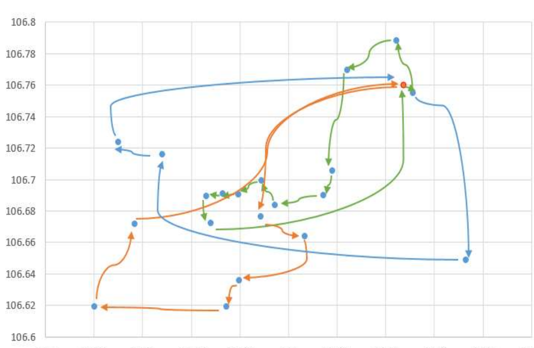


Figure 4. The best route for option 3 (1 big truck & 2 small trucks)

Table 4. Results of options

OPTION	CAPACITY (pallet)	PALLET REMAIN (pallet)	TOTAL DISTANCE (km)	AVERAGE TRUCK UTILIZATION
3 small trucks	75	6	171.8	92%
2 big trucks	72	3	140.9	96%
1 big truck – 2 small trucks	86	17	175.4	77%

As shown in Table 4, the first and second columns indicate the option characteristics while the third column shows the number of pallets remaining between the total vehicle capacity and customers' demands. The fourth column shows the total distances of options while the fifth column shows the average vehicle utilizations. The total distance in option 2 (2 big trucks) is 140.9 km, which is shorter than 22% and 24% compare to the total distance of option 1 (2 small trucks) and option 3 (1 big truck – 2 small trucks). The average truck utilization is the highest in option 2, which is 96%. The average truck utilization in option 3 is 77%, which is less than that of option 1 (92%) and option 2 (96%). From the results, the company can be choosing 2 big trucks to deliver goods from the warehouse to the retailer stores with the inclusion of minimizing total distance and utilizing vehicle capacity.

5. Conclusion

This paper studies the vehicle routing problem by developing a mathematical model. The nearest neighbor procedure is proposed to acquire a good initial solution by minimizing total transportation distance. Based on the real case study, the company can reduce transportation costs for delivery goods with efficient routing.

For future research, the model can be extended by considering vehicle utilization. The transportation cost can be included in the objective function of the model. Besides, enterprises should not only care about distance but also related to many other factors such as cost, time, congestion, infrastructure, etc.

Reference

Agarwal, Y., Mathur, K., Salkin, H.M. (1989) A set-partitioning-based exact algorithm for the vehicle routing problem. *Networks*, 19(7), 731-749.

Bautista, J., Fernández, E. and Pereira, J. (2008) Solving an urban waste collection problem using ants' heuristics. *Computers & Operations Research*, 35(9), 3020-3033.

Balinski, M.L., Quandt, R.E. (1964) On an integer program for a delivery program. *Operations Research*,12, 300-304.

Boussaïd, I., Lepagnot, J. and Siarry, P. (2013) A survey on optimization metaheuristics. *Information sciences*, 237, 82-117.

Bräysy, O., Nakari, P., Dullaert, W. and Neittaanmäki, P. (2009) An optimization approach for communal home meal delivery service: A case study. *Journal of Computational and Applied Mathematics*, 232(1), 46-53.

Clarke, G. and Wright, J.W. (1964) Scheduling of vehicles from a central depot to a number of delivery points. *Operations research,* 12(4), 568-581.

Dantzig, G. B. and Ramser, J. H. (1959) The Truck Dispatching Problem. *Management Science*, 6 (1), 80-91.

Elshaer, R. and Awad, H., 2020. A taxonomic review of metaheuristic algorithms for solving the vehicle routing problem and its variants. *Computers & Industrial Engineering*, 140, 106242.

Gendreau, M., Hertz, A. and Laporte, G. (1994) A tabu search heuristic for the vehicle routing problem. *Management science*, 40(10), 1276-1290.

Gillett, B.E. and Miller, L.R., (1974) A heuristic algorithm for the vehicle-dispatch problem. *Operations research*, 22(2), 340-349.

Laporte, G., Nobert, Y. and Desrochers, M. (1985) Optimal routing under capacity and distance restrictions. *Operations research*, 33(5), 1050-1073.

Laporte, G., Nobert, Y. and Taillefer, S. (1988) Solving a family of multi-depot vehicle routing and location-routing problems. *Transportation science*, 22(3), 161-172.

Laporte, G., Gendreau, M., Potvin, J.Y. and Semet, F. (2000) Classical and modern heuristics for the vehicle routing problem. *International transactions in operational research*, 7(4-5), 285-300.

Oppen, J. and Løkketangen, A. (2008) A tabu search approach for the livestock collection problem. *Computers & Operations Research*, 35(10), 3213-3229.

Pei-Ying, Y.A.N.G., Jia-Fu, T.A.N.G., Yang, Y.U. and Jin-Xiang, P.E.I. (2013) Minimizing Carbon Emissions through Vehicle Routing and Scheduling in the Shuttle Service of Picking up and Delivering Customers to the Airport. *Acta Automatica Sinica*, 39(4),424-432.

Pollaris, H., Braekers, K., Caris, A., Janssens, G.K. and Limbourg, S. (2015) Vehicle routing problems with loading constraints: state-of-the-art and future directions. *OR Spectrum*, 37(2), 297-330.

Renaud, J., Boctor, F.F. and Laporte, G. (1996) An improved petal heuristic for the vehicle routeing problem. *Journal of the operational Research Society*, 47(2), 329-336.

Yildiz, H., Johnson, M.P. and Roehrig, S. (2013) Planning for meals-on-wheels: algorithms and application. *Journal of the Operational Research Society*, 64(10),1540-1550.

PERISHABLE INTERMODAL MULTIPLE PRODUCT SUPPLY CHAIN NETWORKS UNCENTAINTY ANALYSIS BY DISCRETE-EVENT SIMULATION APPROACH: A CASE STUDY OF FRESH FRUIT SUPPLY CHAIN

Luong-Nhat Nhieu[1*], Kha-Hoang Nguyen[1] and Chia-Nan Wang[1]

[1] National Kaohsiung University of Science and Technology, Kaohsiung 80778,

nnluong.iem@gmail.com (L.-N.N.)

cn.wang@nkust.edu.tw (C.-N.W.)

nhkhahn@gmail.com (K.-H.N.)

Abstract. Supply chains designed for perishable products not only need to ensure the basic indicators of sustainable development such as economic, environmental, and social, but also need to be accurate in terms of time as well as high sensitivity to uncertainties. For that purpose, this study develops a simulation procedure that allows analysis of the influence of uncertainties on the efficiency of the entire supply chain. The simulation model is developed based on the supply chain configuration previously designed by a mixed-integer programming optimization model. According to the design of the optimization model, the conceptual models and the simulation models and their validation procedures are performed. In the next step, simulation scenarios are set up with fluctuations in uncertainties such as demand, productivity and transit time. Through comparing and evaluating the output of simulation runs, the author proposes recommendations for supply chain managers to be able to react quickly when external conditions change.

Keywords: Supply Chain; Simulation; Perishable Product; Sustainable Development; Uncertainty Analysis.

1. Introduction

In recent decades, simulation and optimization have been an increasingly popular combination for enhancing and verifying solutions. Accordingly, the role of optimization algorithms is to find possible solutions in a system of complex constraints that achieve the highest efficiency for the objective functions. Meanwhile, simulation methods are often used afterwards to verify the optimal solution in a chaotic environment of uncertainty. Contributions from improving the efficiency of supply chains have led to positive developments for related industries. Supply chain design and management problems involve such complex constraints and uncertainties.

While a study to be published in 2021 has developed a deterministic multi-objective optimization mathematical model aimed at supporting decisions in the design and operation of the perishable product supply chain and applied it for Vietnam Fresh Fruit Case Study (Wang, Nhieu, Chung, & Pham, 2021). The supply chain that this model concerns has 4 tiers including farm, local collection hub, crossdocking center (CC) and consumer market. Their study has also developed and evaluated scenarios, which have an impact on the optimal solution, based on the subjective views of decision makers in terms of the weights of the objective function. Therefore, the missing piece to complete these results are assessments of the uncertainty that the above fruit supply chain may face.

Therefore, the main contribution of this paper is the development and validation of a discrete event simulation model that accurately describes the results from the aforementioned study. This simulation model, which has been validated, is then used to analyze uncertain scenarios with uncertain inputs as random distributions. Based on the results, practical recommendations are proposed for supply chain managers on strategic decisions in an uncertainty environment.

The structure of this paper's remaining part includes a review of related studies in section 2, a description of the methodology in section 3, and presentation of simulation results in section 4. Finally, the article This paper closes with the conclusions in section 5.

2. Literature review

In the Industry 4.0 era, real-time planning in production control is essential, and there is a high chance of optimizing the entire supply chain. In the Patrick's paper, the author has applied it in the factory industry, especially for outsourcing and on-site installation. Old-fashioned planning often ignoring deviations onsite and the overall schedule is not up to date. That leads to reduced productivity, high inventory levels. The author has shown the on-site inventory level through simulation, so that production can be planned more reasonably (Patrick, A., Erwin, & T., 2017). Simulation is a widely used method in the world. In 2017, in order to modeling the system of supply chain processes, W. K. V. Chan et al. used Bayesian approach, but it still has many limitations. However, in some uncertain cases, the Bayesian modeling brings key advantages. To apply this model, the author has combined two methods, Bayesian model with simulation, into the supply chain of manufacturers and distributors. The data source was collected over a specific period of time. To develop this model, it is necessary to understand the product production line well, to manage well the balance of inventory, profit and service (Chan et al., 2017). In supply chain management, structure and process is one of the fundamental challenges. In the research introduced by Dmitry, resilience has several intersections with supply chain sustainability. The purpose of the paper is to analyze supply chain disruptions and the structure of sustainability factors in the supply chain. Resilience in the supply chain involves minimizing ripple effects and increasing sustainability. As a result, the author has shown that individual supplies increase the ripple effect; strengthening facilities reduces the ripple effect and enhances sustainability and vice versa (Dmitry, 2017). In the other supply chain literature, there are many models to approach and apply in practice. In particular, the model of optimizing the supply chain network design has been paid more attention. The methods of linear programming and discrete event simulation are commonly used. By combining analysis models and discrete event simulations to support network design decisions. The purpose of the Chiadamrong's paper is to find out the differences between the solutions and satisfy the given criteria. The results from a conventional simulation model show that the optimal timing is much better than the results obtained. The effect of this combination has great potential for development, providing many solutions to apply in practice (Chiadamrong & Piyathanavong, 2017).

In the study which published in 2018, C. J. Grandzol and J. R. Grandzol mentioned that the supply chain design and constraint management is the process to identify the right balance between inventory, transportation, and manufacturing cost. These two methods are widely adopted techniques in industry. To get special attention from students, it requires that operations and schools to operate and teach strongly on these topics to maximize resources, especially time. To achieve those criteria, the Chantey Castings simulation provides a full platform for students to develop. In a collaborative student-manager environment, students can have fun learning the key concepts of these techniques through hands-on experience with customer needs, sourcing, production and logistics for a variety of products. In this paper, it is shown that simulation is a highly effective development and experiential approach (Grandzol & Grandzol, 2018). As the mentioned statement by Rabe et al, the production line in the workshop is not properly arranged, leading to the waste of time for workers to travel, and is the main cause of the decrease in productivity. To reduce travel distances, consolidated product flows and better resource utilization. Horizontal collaboration is an innovative strategy for product consolidation. Goods Transportation in urban environment, also known as City Logistics, must pay special attention to, cooperation between parties at the same level can ensure efficiency and sustainability. The purpose of this paper is to evaluate the business opportunities of the commercial sectors, as well as the cooperation of transport activities through discrete event simulation model. The data of this paper provided by Greek logistics operators are real and the results show the cooperation of the two approaches regarding urban consolidation (Rabe, Klueter & Wuttke, 2018). The paper by Rossetti and Bright considered the petroleum is one of the indispensable raw materials in society, the establishment of a petroleum supply chain is also very important. This paper describes the petroleum supply chain through simulation. The purpose of the paper is to focus on the analysis process and the petroleum transport capacity. The methodology used is extremely diverse, with a deep understanding of the petroleum supply chain and analytical capabilities to facilitate supply chain. The results show that it is possible to analyze the resilience of the commercial bulk petroleum supply chain under conditions of disruption (Rossetti & Bright, 2018). Besides that, the risk is problem always to the supply chain. Effective risk management leads to better supply chain operations. The intention of author research is to analyse the role and contribution of simulation and optimization methods for the supply chain risk management. Consequently, a faster approach to risk management phases, simulation and optimization methods is desired. A direct consequence of this approach is that the revealed a key methodological disconnection. Although it is a known issue, it has to be highlighted the limitations of several models and complexity of the risks in supply chains, particularly on real-world/real-time applications. The method provides flexibility to handle very complex issues found in supply chains. This suggests that these future developments are intended suggested aiming to develop new applications focused on simulation and optimization tools for risk mitigation proposes (B.Oliveira, Jin, Lima, Kobza, & Montevechi, 2019). In addition, S. Barat's work in 2019 is the one of the well-controlled studies on the control applications such as online gameplay and robotics. However, this method is rarely used for managing operations of business-critical systems such as supply chains. Despite system complexity, it can be accurately modeled by using Reinforcement Learning (RL). In this paper, in order to implement the RL method into a real system, the author has to describe a framework for integrating cybernetic and simulation into a complex networked system (Barat et al., 2019). In another study, Mohammad et al compared the two methods in the blood product delivery system, the method integrated inventory blood supply chain (BSC) and the current method. Blood is the source of life for the body, the importance of blood to human life is unlimited and as are the valuable blood products of BSC. They applied discrete event simulation and discussed the problem to control the complexity of the system. The author using arena simulation package, to model blood supply chain systems to assess important factors in BSC. Although the method is new, if proposed it would be help for future research purposes (Mohammad, Xian, & Saeed, 2020).

3. Methodology

As depicted in Figure 1, the methodology of this paper starts with defining the problem and its properties. Then, data collection and conceptual model development that describe the operation of the fresh fruit supply chain were carried out in parallel. With the information from the data, the conceptual model is transformed into a simulation model in the foundation of Rockwell ARENA discrete event simulation software.

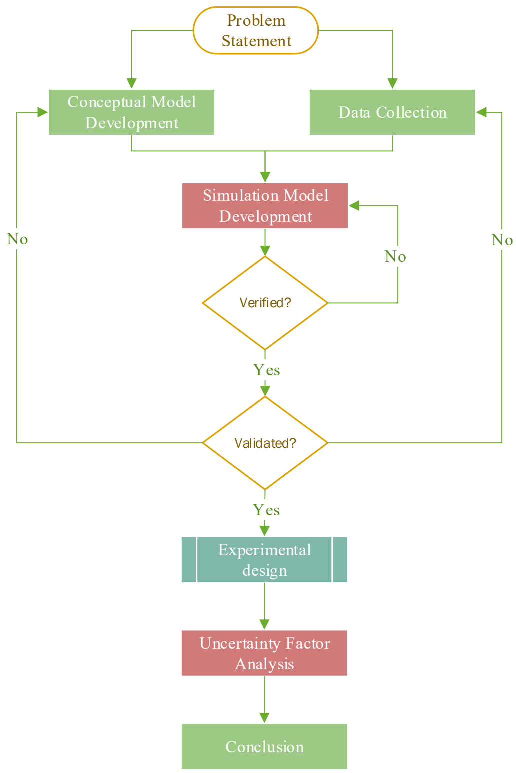


Figure 1. Methodology flow chart

The next two steps verification and validation are very important to ensure that the simulation model not only accurately describes the nature, but also gives an accurate result of the output compared to the optimization model's results in the parent research. Then, uncertainty tests are designed for simulation runs. The results of these runs are analyzed to support judgments about the impact of uncertainties on supply chain configurations. Finally, on the basis of the above conclusions, this study proposes recommendations on strategic supply chain management decisions for decision makers.

3.1. Conceptual model

The conceptual model, which is depicted in Figure 2, represents the product flows and decision logics of the supply chain. In it, routes and transport quantities are defined and assigned based on the optimization results. The mode of transport in this supply chain includes inland waterways and roads which have a predetermined proportion in each segment of the supply chain according to the optimized solution. The speed and load of the vehicle depends on its design parameters. However, in the simulation model, these parameters can be identified as random distributions.

3.2. Data collection

The data used in this simulation model is based on the deterministic parameters of the optimization model. Accordingly, this study converts those deterministic parameters into uncertain inputs as triangular random distributions. Where, the deterministic value is used as the mean of the triangular distribution inputs. Meanwhile, the upper and lower bounds are determined through a deviation coefficient (φ) from the mean. This implies that the larger the value of the coefficient φ, the larger the amplitude of the random inputs. As an example, the deterministic packing processing time (PPT) of each worker at the crossdocking center is m tons/hour. Thus, the packing processing time in the simulation model follows the distribution TRIA (a, b, c), where the value of a, b, and a are determined by equations (1), (2), and (3).

$$a = m \times (1 - \varphi) \quad (1)$$

$$b = m \quad (2)$$

$$c = m \times (1 + \varphi) \quad (3)$$

3.3. Simulation model

Based on the conceptual model and the above data, this study develops a simulation model that consists of four submodules corresponding to the four tiers of the supply chain. Along with that are global variables to measure the objective functions throughout the runs.

3.3.1. Simulation model validation

For the purpose of validating the model before performing further analysis, this study performed the first simulation with thirty replications, in which each replication was 365 days in length. Then, the outputs obtained from the simulation are compared with the corresponding values from the optimization study to determine the compatibility between the simulation model and the optimization results (μ_0). This comparison is performed at three outputs, which are the total number of products entering the supply chain in farms, the total number of products leaving the supply chain in markets and the values of the objective functions via confidence interval testing (CI testing). The test results are presented in Tables 1, 2 and 3.

The confidence interval testing procedure is started by determining the confidence interval of the outputs with a confidence level of 5% ($\alpha = 0.05$). As shown in Figure 2, based on whether the optimization value is covered by the confidence interval of the outputs determines the Best-Case (BCV) and Worst-Case (WCV) values of the test.

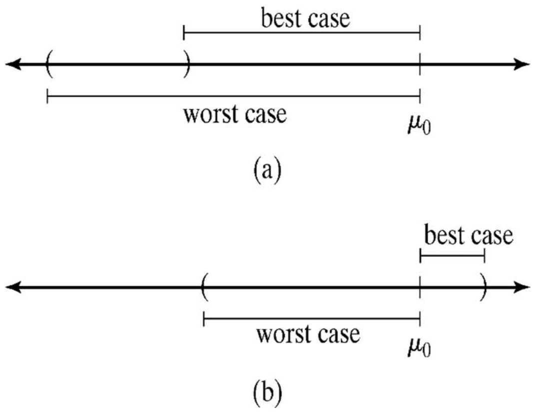


Figure 3. CI testing Best-Case and Worst-Case value. (a) CI does not contain μ_0. (b) CI contains μ_0.

In the next step, the best-case and worst-case values are compared with a given acceptable error (ε) that is chosen to be 5% of the μ_0. In the case where the CI does not contain μ_0, the simulation model needs to be adjusted if $BCV > \varepsilon$ or is considered validated if $WCV \leq \varepsilon$ or needs more replications to evaluate if $BCV \leq \varepsilon$. In the case where the CI contains μ_0, the simulation model is considered validated if $WCV \leq \varepsilon$. The model needs more replications to evaluate if WCV or $BCV \geq \varepsilon$. Accordingly, the validation process results show that the percentage of outputs that are validated, need more replications and need to be adjusted are 84.62%, 11.54% and 3.85% respectively. Therefore, this simulation model is evaluated as representing the correct performance of the supply chain configuration that is determined by the optimization solution. Thus, the simulation model developed in this study can be used to examine the effects of uncertainty on the performance of this supply chain through adjustment for the deviation coefficient (φ) introduced.

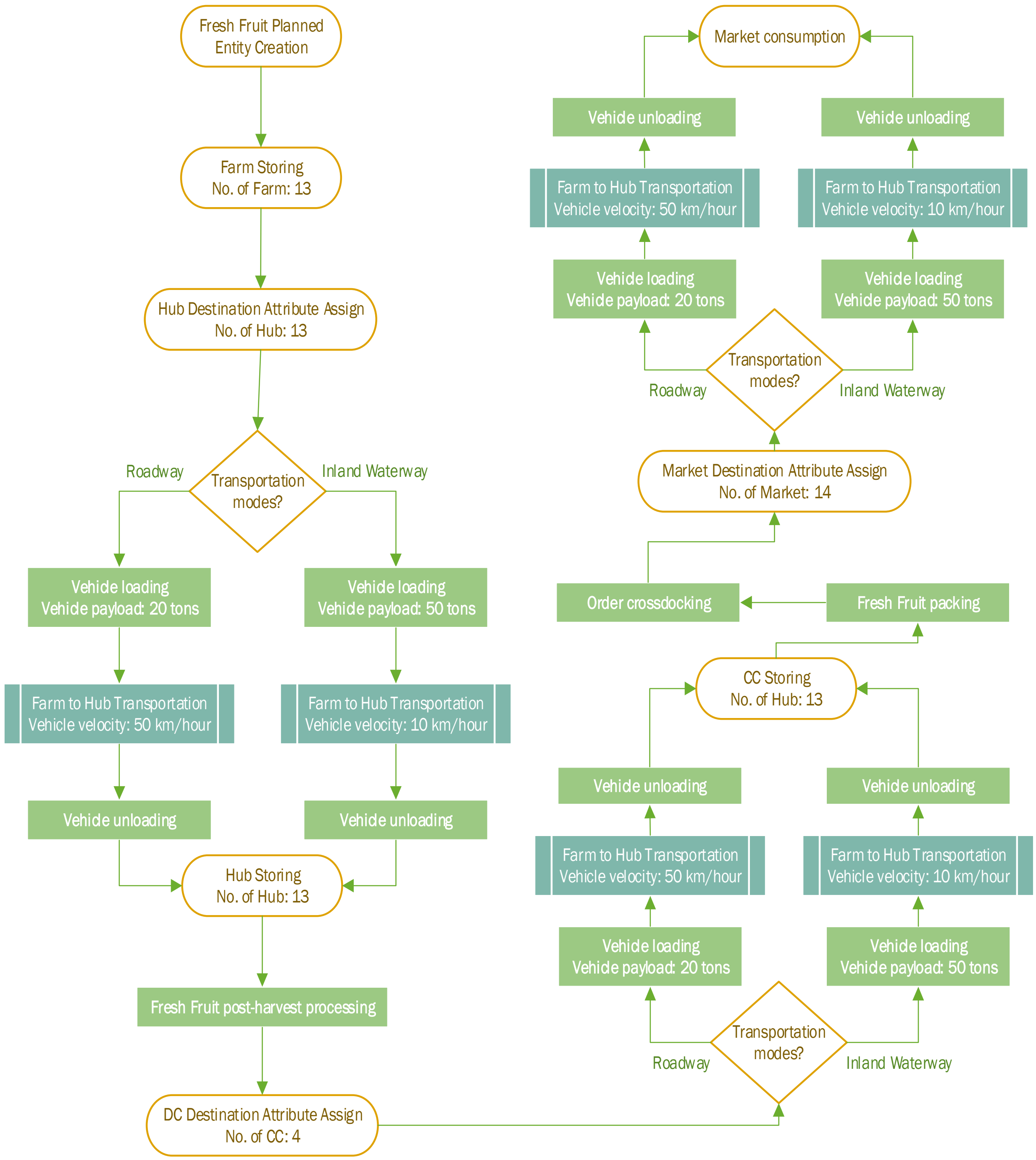


Figure 2. Conceptual model

Table 7. Fruit entity entering model

Product	Optimization value	Simulation Output		Confidence Interval	Best Case	Worst Case	Acceptable Error	Validation results
		Mean	Half Width					
Dragon fruit	583,900	563,989	1,588	[562401;565577]	18,323	21,499	29,195	Validated
Mango	252,800	246,624	3,241	[243383;249865]	2,935	9,417	12,640	Validated
Rambutan	27,500	29,461	904	[28557;30365]	1,057	2,865	1,375	More Replication
Durian	210,000	212,193	905	[211288;213098]	1,288	3,098	10,500	Validated
Star apple	70,000	71,289	2,183	[69106;73472]	894	3,472	3,500	Validated
Pomelo	500,000	487,580	1,567	[486013;489147]	10,853	13,987	25,000	Validated
Longan	210,397	222,484	3,947	[218537;226431]	8,140	16,034	10,520	More Replication
Banana	1,283,220	1,244,976	1,767	[1243209;1246743]	36,477	40,011	64,161	Validated
Pineapple	524,000	898,956	4,019	[894937;902975]	370,937	378,975	26,200	Refine Needed
Orange	472,500	468,165	2,273	[465892;470438]	2,062	6,608	23,625	Validated
Mandarin	36,750	37,434	845	[36589;38279]	161	1,529	1,838	Validated

Table 8. Fruit entity leaving model

Product	Optimization value	Simulation Output		Confidence Interval	Best Case	Worst Case	Acceptable Error	Validation results
		Mean	Half Width					
Dragon fruit	578,946	559,250	1,588	[557662;560838]	18,108	21,284	28,947	Validated
Mango	249,502	242,293	3,241	[239052;245534]	3,968	10,450	12,475	Validated
Rambutan	24,355	25,467	904	[24563;26371]	208	2,016	1,218	More Replication
Durian	207,867	208,278	905	[207373;209183]	494	1,316	10,393	Validated
Star apple	66,719	67,231	2,183	[65048;69414]	1,671	2,695	3,336	Validated
Pomelo	496,948	483,443	1,567	[481876;485010]	11,938	15,072	24,847	Validated
Longan	206,636	202,491	3,947	[198544;206438]	198	8,092	10,332	Validated
Banana	1,280,518	1,241,393	1,767	[1239626;1243160]	37,358	40,892	64,026	Validated
Pineapple	521,993	516,017	4,019	[511998;520036]	1,957	9,995	26,100	Validated
Orange	469,123	464,284	2,273	[462011;466557]	2,566	7,112	23,456	Validated
Mandarin	34,556	34,835	845	[33990;35680]	566	1,124	1,728	Validated

Table 9. Objective functions

Objective	Optimization value	Simulation Output		Confidence Interval	Best Case	Worst Case	Acceptable Error	Validation results
		Mean	Half Width					
Transportation Cost (USD)	95,421,270	94,930,377	118,119	[94812257;94812257]	609,013	609,013	4,771,063	Validated
Transportation Time (Hour)	606,547	625,558	800	[624758;624758]	18,211	18,211	30,327	Validated
Transportation Emission (Kg CO2)	35,547,459	34,315,762	37,452	[34278309;34278309]	1,269,150	1,269,150	1,777,373	Validated
No. of Mismatching (Tons)	53,101	54,871	129	[54742;54742]	1,641	1,641	2,655	Validated

3.3.2. Simulation experiments

In this section, simulation experiments are built based on the magnitude of the deviation coefficient (φ) that imply for the magnitude of the uncertainties as Table 4.

Table 10. Simulation experiments

Experiment	Deviation coefficient value(φ)	No. of Replications
1	0	30
2	0.1	30
3	0.2	30
4	0.3	30

The results of the empirical analysis show that increasing the width of the triangular random distributions leads to unpredictable changes in the objectives' replication mean as presented in figure 4, 5 ,6, and 7. However, the distances between the maximum and minimum values between simulation replications increased with the increase of the deviation coefficient φ. This implies that the intensity of uncertainties at individual points at the supply chain tiers can lead to large fluctuations in the efficiency of the entire chain. Therefore, for the situation not to become too unpredictable, supply chain managers need to be concerned with standardized solutions that work at every single point across the supply chain.

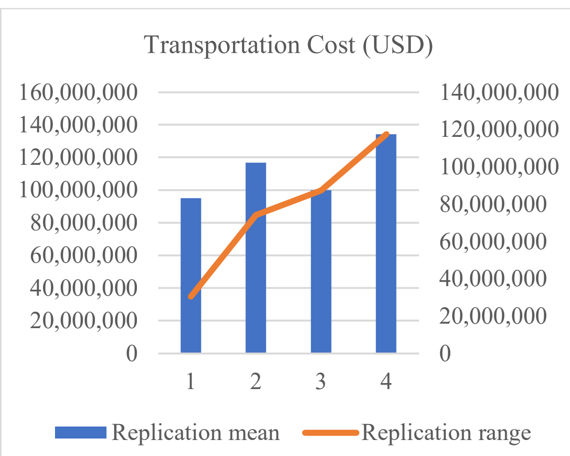


Figure 4. Transportation cost replication mean and range

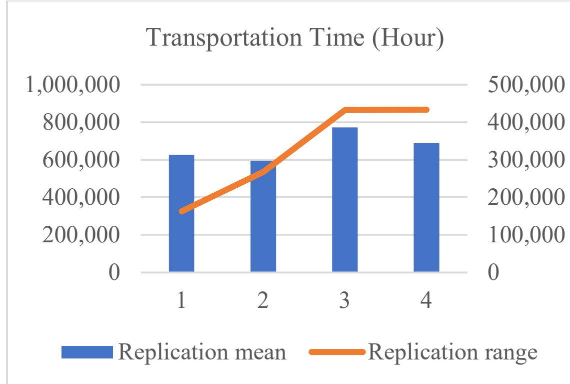


Figure 5. Transportation time replication mean and range

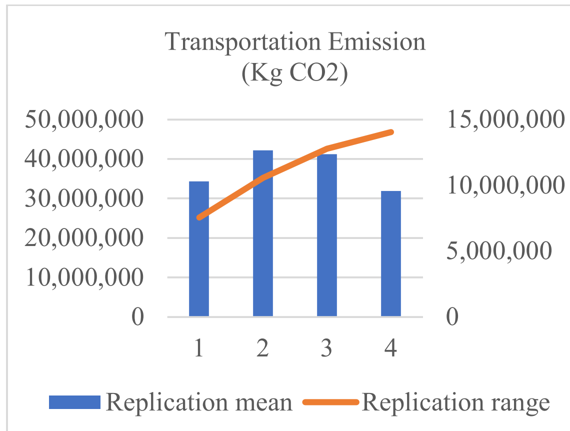


Figure 6. Transportation emission replication mean and range

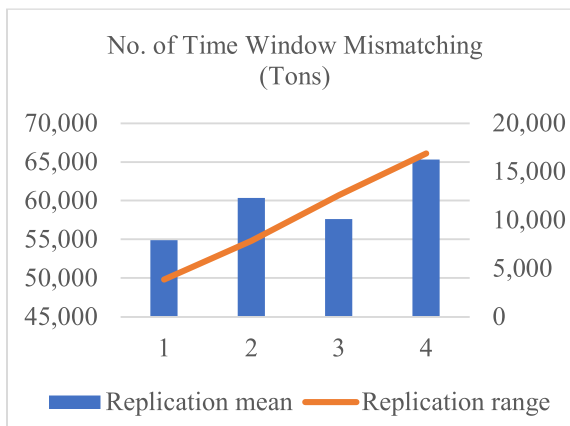


Figure 7. Earlier/later time window delivery replication mean and range

Another finding from the simulation results of this study is that the utilization of human resources at the facilities is relatively low as Figure 8 and Figure 9. The optimization results helped determine the ceiling of the workforce level at the facilities. However, simulations show that with product flow rates into facilities and processing rates as determined, workforce levels can be adjusted down to save costs and increase utilization.

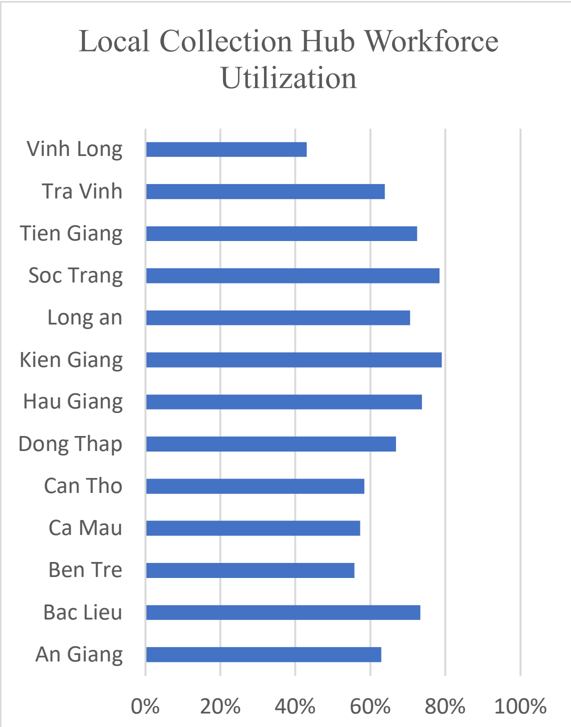


Figure 8. Earlier/later time window delivery replication mean and range

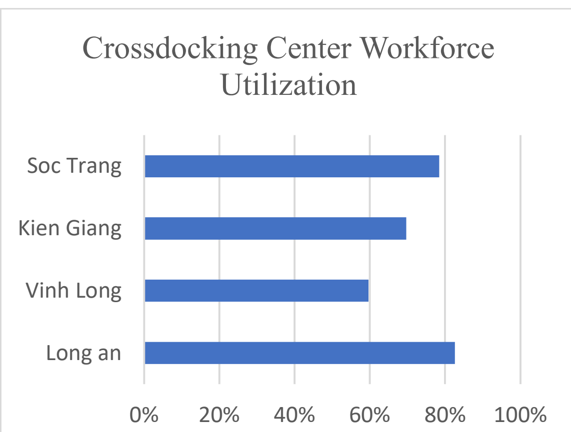


Figure 9. Earlier/later time window delivery replication mean and range

Workforce level adjustments will continue to be made through simulation and further analysis of this study.

4. Conclusion

This study was conducted to take advantage of the simulation approach for verifying and enhancing the solutions from the optimization models. Accordingly, based on the optimization results, a four-tier supply chain simulation model is built and validated. The simulation results show the changing effects and trends of supply chain performance in an uncertain environment. The deterministic parameters and deterministic solutions of the optimization model such as the workforce level can also be simulated and improved in the next phase of this study.

Reference

B.Oliveira, J., Jin, M., Lima, R. S., Kobza, J. E., & Montevechi, J. A. B. (2019). The role of simulation and optimization methods in supply chain risk management: Performance and review standpoints. *Simulation Modelling Practice and Theory, 92*, 17-44. doi:10.1016/j.simpat.2018.11.007

Barat, S., Khadilkar, H., Meisheri, H., Kulkarni, V., Baniwal, V., Kumar, P., & Gajrani, M. (2019). Actor Based Simulation for Closed Loop Control of Supply Chain using Reinforcement Learning. *International Conference on Autonomous Agents and Multiagent Systems, 18*, 1802-1804.

Chan, W. K. V., D'Ambrogio, A., Zacharewicz, G., Mustafee, N., Wainer, G., & Page, E. (2017). A Bayesian Simulation Approach for Supply Chain Synchronization. *Winter Simulation Conference*.

Chiadamrong, N., & Piyathanavong, V. (2017). Optimal design of supply chain network under uncertainty environment using hybrid analytical and simulation modeling approach. *Journal of Industrial Engineering International, 13*(4), 465-478. doi:10.1007/s40092-017-0201-2

Dmitry, I. (2017). Revealing interfaces of supply chain resilience and sustainability: a simulation study. *International Journal of Production Research, 56*(10), 3507-3523. doi:10.1080/00207543.2017.1343507

Grandzol, C. J., & Grandzol, J. R. (2018). Chantey Castings: A Hands-On Simulation to Teach Constraint Management and Demand-Driven Supply Chain Approaches. *Decision Sciences Journal ofInnovative Education, 16*(1), 6-22.

Mohammad, A., Xian, L., & Saeed, A. (2020). Scenario-Based Simulation Approach for An Integrated Inventory Blood Supply Chain System. *Winter Simulation Conference*, 1348-1359. doi:10.1109/wsc48552.2020.9384018

Patrick, D., A., R. R., Erwin, R., & T., M. D. (2017). Simulation Based Validation of Supply Chain Effects through ICT enabled Real-time-capability in ETO Production Planning. *Procedia Manufacturing, 11*, 846-853. doi:10.1016/j.promfg.2017.07.187

Rabe, M., Klueter, A., & Wuttke, A. (2018). Evaluating The Consolidation of Distribution

Flows Using a Discrete Event Supply Chain Simulation Tool: Application to a Case Study in Greece. *Winter Simulation Conference*.

Rossetti, M. D., & Bright, J. (2018). Bulk petrolium supply chain simulation modeling. *Winter Simulation Conference*.

Wang, C.-N., Nhieu, N.-L., Chung, Y.-C., & Pham, H.-T. (2021). Multi-Objective Optimization Models for Sustainable Perishable Intermodal Multi-Product Networks with Delivery Time Window. *Mathematics, 9*(4), 379. doi:10.3390/math9040379

PREFERENTIAL POLICIES IN ATTRACTING FOREIGN DIRECT INVESTMENT IN VIET NAM

Le Ngoc Thom[1]

[1] *University of Finance and Business Admisstration, Email: ngocthomlee@gmail.com*

Abstract. This paper studies the impact of preferential policies on attracting foreign direct investment (FDI) in Vietnam from 2010 to 2019. By means of meta-analysis, descriptive statistics, collecting data posts have pointed out the influences of financial policies such as tax incentives, land policies, import and export duties ... during last stage. From there, assessed financial policies according to 4E model (Effectiveness, Efficiency, Economy, Equality) pointed out successes and limitations of preferential policies in FDI attraction activities. Then, there are some orientations to attract FDI in the next period.

Keywords: FDI, preferential policies, FDI attraction policy.

1. Introduction

Foreign direct investment (FDI) plays an important role in socio-economic development in the innovation process in developing countries. In Vietnam, FDI not only brings a large amount of investment capital, contributes to GDP growth, affects the process of economic restructuring, and improves social security issues such as creating jobs for people, but also promoting foreign trade activities such as import and export as well as bringing the spillover effects of technology or managerment skills into the host country. At the 30-year FDI attraction conference to be held on October 4, 2018 at the National Convention Center, the government affirmed "FDI is an important driving force of the economy and attracting FDI. is a part of the national economic development strategy in general, sectors and localities in particular".

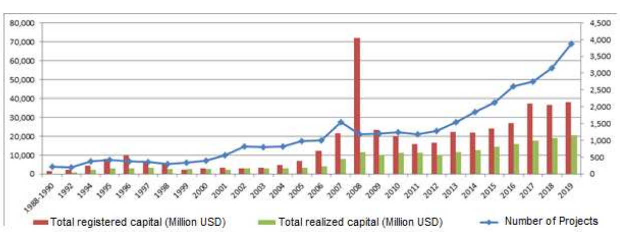


Figure 1: FDI growth rate in the period 1988 – 2019

In the context of a developed economy with many uncertainties as at present, the competition between regions, countries and localities has become increasingly fierce to attract foreign direct investment. In this competition, attracting investors needs more attention and creates a sustainable attraction. In the guiding document of the Politburo's Resolution No. 50-NQ / TW dated August 20, 2019, on the orientation to perfect institutions, policies, and improve the quality and efficiency of foreign investment cooperation to 2030 gave the point of view: "Actively attracting and cooperating with foreign investment selectively, taking quality, efficiency, technology and environmental protection are the main evaluation criteria. Prioritize projects with advanced technology, new technology, high technology, clean technology, modern governance, high added value, spillover effects, connecting production and supply chains. bridge." Therefore, investment attraction policies need to be improved in order to create an environment to attract FDI more effectively.

2. Theoretical framework

2.1. Attracting foreign direct investment (FDI) capital

Attracting foreign direct investment is a concern of all economies, especially in developing countries like Vietnam. Research on FDI not only attracts researchers but also the interest of policy makers.

In The Theory of Marginal Returns, Mac. Dougall proposed a theoretical model that the capital flow of investment will shift from a country with low interest rates to a country with high interest rates until an equilibrium is reached (the interest rates of the two countries are equal) (1960). After investment, the two countries above are profitable and cause the general output of the world to increase compared to before the investment. However, this theory does not explain the phenomenon why some countries simultaneously capital inflows have, capital outflows; There is no full explanation for FDI. Therefore, the theory of marginal returns can only be considered as the first step to studying FDI.

Industrial organization theories (Industrial organization theories) were born in the early 1960s before the formation of large corporations, transnational investment in the US and tried to explain this phenomenon. Stephen Hymer said that, due to the structure of the monopoly market, the US companies to expand their branches abroad to exploit the advantages of capital, technology, management techniques, and market networks. companies in the same industry in the host country are not available. That is the cause of the formation of MNCs and the establishment of branches abroad. Robert Z. Aliber explains the phenomenon of FDI from the effect of tax factors and market size on monopolies. According to Z. Aliber, taxes have increased import prices, so companies have to move production abroad to overcome protective tariff barriers to reduce costs, lower costs, and increase profits. On the other hand, the economic efficiency depends on the size of the market, so monopolies have expanded the market by establishing branches abroad. Products made with new technology tend to be exclusive due to their low cost, so they actively expand their production overseas to take advantage of exclusive rights, Richard E. Caver explains. techniques to maximize profits, thereby forming FDI.

Some studies have demonstrated the influence of institutions on foreign investment location decisions. These include a meta-analysis of institutional influence at the country and sectoral levels (Bailey, 2018), and identify specific determinants of institutional influence on site choice. (Jain, Kothari and Kumar, 2016; Kim & Aguilera, 2015), as well as overviews of the position determinants identified from various theories (Jain et al., 2016; Kim & Aguilera, 2015; Nielsen, Asmussen, & Weatherall,2017).

Maria Guadalupe Lugo-Sanchez (2018) in her study "The Role of Public Policies in Attracting Japanese FDI in Mexico" pointed out the role of public policies, including tax policies and tax incentives, in attracting FDI. from Japan to Mexico.

And, in the studies of Aguilera and Grøgaard (2019) and Cuervo-Cazurraet al. (2019) pointed out the role of institutions in general and policies, including preferential policies for foreign direct investment.

In general, studies have pointed out a number of important factors affecting foreign direct investment in the region, territory or locality in terms of attracting factors from the receiving country, of which like element of the policy of FDI, economic factors, favorable business factors: such as investment promotion, investment incentives or hardware infrastructure ...

Of the factors attracting FDI, the coefficient system economic policy related to foreign investment increasingly important role. The policies attract investment of recipient countries need to be very flexible and attractive to become an important condition for collecting FDI such as:

- Policies to encourage foreign direct investment: This is one of the prerequisites when foreign investors are interested in a certain investment location. Resonable incentives policies will create favorable conditions for investors, whereas unresonanle investment policies will create unreasonable barriers to large, creating an investment environment is not favorable to the person first investment.
- Preferential policies on finance: If government want to attact foreign investors investing in the country, in the locality, in priority industries or fields, they must give foreign investors certain incentives and supports, specifically in terms of finance to create the best conditions for investors to make profits.
- Preferential policies on taxation: countries must have certain and reasonable tax exemption policies to attract investment. Normally, in the first years of project implementation, investors are entitled to tax reduction, even tax exemption and gradually increase in the following years when foreign investors make profits.

Besides, there are also several policies such as trade policy, exchange rate policy ... affecting FDI attraction.

2.2. Preferential policies and FDI attraction

Financial policies are a tool regulated the macroeconomic of the State, a part of the socio-economic policy, as "criterion for handling financial distribution relations according to the interests of the ruling class" (Nguyen Van Xo, 1998). Currently there are many ways to classify financial policy, such as the scope of activities with national financial policy, the local financial policy or the movement of financial sources and the policy of using financial resources.

Preferential policies to attract foreign direct investment are part of the National Financial Policy. This is an important macro management policy system, including a system of policies on budget revenue and budget expenditure (investment expenditure, recurrent expenditure); debt, aid; monetary policy - credit, insurance, corporate finance, external finance, savings, residential finance and financial management of national capital and assets. Financial policies to attract FDI include: (i) specific financial mechanisms and policies for development by economic zones or socio-economic regions such as industrial zones and processing zones. Exports, hi-tech zones, open economic zones, special economic zones, etc. Foreign investors investing capital to establish businesses and operate in these regions or areas will benefit or share responsibilities. from separate financial policies applicable to these particular territories; (ii) general financial policy mechanisms applicable to the whole country such as key industries, high technology, agriculture or specific entities (SMEs).

Financial policies have direct impact on attracting of foreign direct investment capital through credit policies, State budget revenues and expenditures ...

Firstly, about state budget expenditure: Policies on spending of the state budget for infrastructure, investment promotion, science and technology innovation, human resource training ... will create a positive environment to attract foreign direct investment. Spending on trade promotion and investment promotion will directly affect Vietnam's image with foreign investors and attract this capital source.

Secondly, about tax policies. Tax policies can orient and regulate the government's point of view with foreign direct investment. Through specific preferential regimes, specific taxes will affect the investment capital and the quality of foreign direct investment.

Thirdly, land policy. Through specific policies on land fund, support for land clearance, exemption of land rent ... for projects are encouraged, which will be an opportunity to attract foreign investment in industries, sectors, local under the direction of the government.

2.3. The 4E model evaluates policy effectiveness

Policy evaluation is usually based on three criterias: effectiveness, efficiency, and economy (also known as 3E model: Efficiency, Effectivness and Economy). However, some countries like the US, in addition to the 3E mentioned above, also add Equality. Effectiveness is assessed on the results achieved by the policies. Efficiency is assessed by the access to preferential policies and the binding of the policies. Economy is evaluated through the cost of time, hard cost and soft cost in preferential policies. Equality: evaluate the equality between domestic and foreign investors.

The table of evaluation of financial policies affecting FDI attraction in Vietnam can be summarized as follows:

Table 1: The 4E Model

Criteria of Evaluation	**Observe reality**
Effectiveness	- FDI enterprises have access to financial incentives. Preferential tax policies have contributed to facilitating agreement beneficial to work up the production of sales of FDI enterprises. - Policy without the constraints of the term: the provisions on income tax incentives for investment projects are absolute time period from when the project went into operation (for preferential tax rate) and since there is taxable income from the investment project (for incentives or reductions). This regulation easily leads investors to take advantage of preferential policies to benefit, after the incentive period expires, investors can invest in new projects in other localities.
Economy	- The cost of applying preferential policies to attract investment in Vietnam is quite large. - Preferential policies on tax exemption and reduction period can encourage businesses to avoid taxes through investment restructuring, into new investment projects to continue enjoying incentives. Besides, the types of profit incentives tend to attract short-term investments, with budget costs often large and often unpredictable. - The integration of social policies into CIT incentives has made the tax policy more complicated, difficult to manage, and has created favorable conditions for price transfer activities. - The regulations on tax incentives in Vietnam are quite complicated, increasing the tax compliance costs.
Efficiency	- Create a favorable environment for attracting the investment, attracted by province, according to the field lines. - Increasing exports and state budget revenues. - Contribute to positive change in investment structure by region - Deals are also spread out and complicated. The CIT Law currently stipulates 3 preferential tax rates based on industries, fields of interest, investment areas, and capital scale in accordance with the Law on Investment. - The effectiveness of financial incentive policies in encouraging investment in underdeveloped regions and priority areas is very limited. - The situation of dodging the law on import and export taxes is quite common.
Equality	- Agree to apply incentives between domestic and foreign investors - The tax incentives are creating inequalities among enterprises when considering the level of access from pros Deals

3. Impacts of preferential policies on attract investment FDI period 2010 - 2019

Over the past years, Vietnam has continuously improved institutions as well as perfecting preferential financial policies to attract and manage better foreign investment resources. During the period from 2010 - 2019, preferential finance policies focus on: (i) Preferential tax corporate income, (ii) Preferential tariffs and (iii) Preferential policies on land finance. The reform of the legal system and policy mechanisms has promoted a more efficient investment market. The preferential policies attracting FDI from now this is carried out under the new law supplemented, amended as: The Law of investment in 2014, the Law Tax Corporate Income (2008 and amended in 2013), the Law on amendments and supplements Supplementing a number of articles of the tax laws 2014, Consolidated Document 14/ VBHN-VPQH merged the Law on Corporate Income Tax, Law on Import and Export Tax 2016, Decree No. 118/2015/ND-CP detailing and directions to guide the implementation of a number of articles of the Investment Law, Decree 123/2017/ND-CP amending a number of articles on collection of land use fees, collection of land rent and water surface rent, Decree 57/2018/ND-CP on mechanisms and policies to encourage enterprises to invest in agriculture and rural areas, Law on non-agricultural land use ... In which, specific policies such as:

3.1. About policy Income Tax Industry

Before 2010, the economy had developed steps after years of attracting FDI and applying the law to encourage domestic investment. Since the 2005 Investment Law has applied the general investment law nationwide. Besides, the international commitments to multilateral and bilateral is done leading to significantly cut revenues from import duties, Vietnam has carried out tax reform stage 3 with the focus of the reform is geared and realizing three goals: simple, fair and effective. It can be said that the tax policy has made an important contribution to encouraging the development of production and business, promoting exports, encouraging enterprises to invest capital in areas with difficult natural conditions to create development. develop evenly among regions of the country.

Tax policies during this period have left important point about regulation of tax liabilities and incentives apply uniformly to all businesses in the country and now foreign investors with the corporate income tax rate 28% for general application (CIT Law 2003). This tax rate is adjusted in the following years until now as the revised Law on Corporate Income Tax 2014. This policy contributes to the elimination of discrimination among economic sectors, creating a fair about tax obligations for all investors. In addition, businesses also enjoy preferential tax-free income for some income from scientific research contracts, from production of experimental products ...

From 2011 up to now, the Government continues to implement tax reform phase 4. During this period, Vietnam was also affected by the world economic crisis, and a number of problems in the exploitation of natural resources, capital and labour attracted to Vietnam's economic growth slow down. The economic context of international integration and in-depth development has posed a problem for Vietnam to grow in the direction of improving quality and ensuring sustainability. Therefore, the tax policy system continues to be reformed for the 4th time with the most important change to increase tax competitiveness and attract investment with the policy of reducing the common corporate income tax rate. Specifically, the common tax rate through the revisions of the Law on Corporate Income Tax is on a downward trend, in the period 2009-2013 is 25%, in the period 2014-2015 is 22% and from January 1, 2016 is 20%. In addition to the tax rate reduction, the Law amending and supplementing a number of articles of the Law on Corporate Income Tax 2013 has added incentives for investment in industrial zones (except for industrial zones in areas with economic conditions. Economic - Social advantages) and expansion investment projects; High-level corporate income tax incentives for a number of key fields that need investment encouragement have contributed to attracting investment, encouraging business, and facilitating enterprises to increase accumulation. In addition, the Law amending and supplementing a number of articles of tax laws No. 71/2014 / QH13, effective from January 1, 2015, has added a number of fields and occupations subject to priority tax incentives in the agricultural sector such as: Cultivation, husbandry, processing of agricultural, forestry and aquatic products (not apply incentives to the field of forest product processing); manufacturing supporting industry products; large-scale and high-tech production projects... Specifically for FDI enterprises:

Firstly, tax incentives: Apply the tax rate of 10% for 15 years and tax exemption for 04 years, 50% reduction of tax payable for the next 9 years for:

(i) Income of enterprises from performing new investment projects in the fields of: Scientific research and technological development; software product manufacturing; production of composite materials, light construction materials and rare and precious materials; production of renewable energy, clean energy, energy from waste destruction; biotechnology development.

(ii) Income of enterprise from implementing new investment projects in the field of environmental protection, including Manufacture of equipment for environmental pollution treatment; Production and application of environmental protection inventions protected by the State in the form of invention patents or utility solution patents; ...

(iii) An enterprise's income from performing a new investment project in the manufacturing sector (except for projects to manufacture goods subject to special consumption tax, mineral exploitation projects) meets one of two criteria. The following: Projects with minimum investment capital of VND 6 trillion, disbursed no more than 03 years after being granted investment licenses and having a minimum total revenue of VND 10 trillion / year at the latest 03 years since the year of revenue. Or d and implement projects with capital to invest at least 06 trillion, the disbursement shall not exceed 03 years from the licensed investors and employs over 3,000 employees at the latest after 03 years from the year the enterprise collection.

Secondly, tax incentives: Apply the tax rate of 17% for 10 years and tax exemption for 02 years and reduce 50% of payable tax amount for the next 04 years, applicable to corporate income from implementation new investment projects in the fields of: High-grade steel production; manufacturing energy-saving products; manufacture of machinery and equipment for agricultural, forestry, fishery and salt production; manufacture of irrigation equipment; production and refining of feed for cattle, poultry and aquatic animals; develop traditional professions.

In addition, there are several other tax incentives: It is allowed to carry forward losses, businesses with losses can carry forward losses to the next year; allowing quick depreciation under specific conditions; preferential corporate income tax reduction in several specific fields to solve social problems; Preferential special consumption tax for the production and consumption of a number of goods such as cars with small capacity, cars…

3.2. Preferential tax about import and export

Since 2007, Vietnam became a member of the WTO, officially joined the global trading system, which has become an important foundation for Vietnam to continue expanding international trade relations, deep and wide integration and join other multilateral and bilateral free trade agreements. As of 2019, the total number of free trade agreements (FTAs) that Vietnam has been implementing commitments to abolish import duty is 11 and inrto the 2020 agreement was 14. All FTAs aim to eliminate or reduce tariff barriers with a high degree, thereby having certain effects on state budget revenues from import and export activities. Therefore, the Law on Export and Import Tax has also been updated and revised in 2005 and 2016 to meet the requirements of integration commitments, while continuing to improve export preferential policies and attracting foreign investment. From 2016 up to now, the preferential policies have been applied under the Law on Export and Import Tax 2016 on the basis of inheriting Law 2005 and have some adjustments in accordance with practical requirements. It continues to be maintained for raw materials and materials invested enter export to part of the goods made for export; machines, equipment, devices being imported to form fixed assets of the project promotion investment promotion; tax exemption for goods temporarily imported for re-export; tax exemption for goods imported exclusively for scientific research, education and training.

From 2016 to now, preferential policies are applied under the Law on Import and Export Tax 2016 on the basis of inheriting Law 2005 and have some adjustments to suit practical requirements. Accordingly, the Law has added high-tech enterprises, science and technology enterprises, and science and technology organizations that are exempt from import tax on raw materials, supplies and components that cannot be domestically produced within 5 years from the start of production; supplementing the regulations on tax exemption for raw materials, supplies and components which cannot be imported domestically for the manufacture or assembly of medical equipment, should be prioritized for research and manufacture.

Some export and import tax incentives are being applied to foreign investors such as : (i) Import duty exemption for goods imported for processing for foreign countries and when exporting products to the foreign countries are exempt from export tax; (ii) Goods imported for processing are exempt from tax, goods temporarily imported for re-export and goods that are raw materials for the production of exports may be extended the tax payment period up to 275 days from the date of open customs declarations; Temporarily imported and re-exported goods may be extended for tax payment up to 15 days from the date of expiration; (iii) Exemption from import tax on goods to create fixed assets for projects investing in areas of special investment incentives, encouraged investment fields and investment projects in the locality. have difficult socio-economic conditions ...

In general, the reduction of corporate income tax rates and diversification of tax incentives have contributed to creating a favorable investment environment to attract foreign investment.

3.3. Preferential policies on land finance

In the period from 2010 to 2013, financial incentives on land were implemented in accordance with the provisions of the 2003 Land Law and from 2014 up to now under the revised Land Law 2013. Accordingly, from before from 2010 to June 30, 2014, land incentives were divided into two forms of land allocation with collection of land use fees and land lease, with a reduction of the payable amount of 20% / 30% / 50% or exemption for a period of 7 years / 11 years / 15 years. From the July 2014 up to now, the land incentives have complied with the Law of Land 2013 and the Decrees regulating the collection of land use fees, collection of land rents and water surface rents.

Some incentives in this period are: Exemption from land rent, water surface rent for investment projects in fields or geographical areas eligible for investment incentives in accordance with the investment law; Incentives in the field of socialization; Incentives when investing in the agricultural and rural sectors; Expanding the application of the land price adjustment coefficient method and the subjects of application to determine financial obligations on land rent for enterprises that have reformed administrative procedures, shortened the time of determination and notification, pay land rent into the state budget and partially support businesses ... At the same time, to attract investment, strengthen management, and use effectively land financial resources in economic zones, High-tech Zone; The Government has issued Decree No. 35/2017 / ND-CP dated April 3, 2017 regulating the collection of land use fees, land rental and water surface rent in economic zones and hi-tech zones with incentives. is higher than the preferential rate of conventional investment projects.

In addition, to support businesses, the Government has issued many policies and supports on land such as: (i) 50% reduction of land rent in the period 2011-2014; (ii) Adjusting to reduce the rate (%) of common land rental calculation from 1.5% (specified in Decree No. 121/2010 / ND-CP) to 1% (regulated in Decree No. 46 / 2014 / ND-CP) and Provincial People's Committees specify the rate (%) in the frame from 0.5% to 3% for each area, route corresponding to each land use purpose to apply local land rent collection; (iii) Provisions on the application of land price adjustment coefficients in determining land prices to calculate and collect land rent for land plots or land in which the value of the area for land rental calculation is based on the land price in the land price table VND 30 billion for centrally-run cities; under 10 billion.

3.4. Others

In addition to tax incentives and land policies, the government also offers several other incentives for foreign direct investment such as:

Incentives for small and medium enterprises (SMEs): Law on SMEs support 2017 Small and medium enterprise development was born, creating conditions for FDI enterprises to enjoy more preferential policies. In addition, Decree No. 34, 38, 39/2018 / ND-CP has supported the development of microenterprises, SMEs focus on 08 groups including: Access to credit; SME Credit Guarantee Fund; Tax and accounting support; Production premises; Technologies, incubators, technical establishments, co-working areas; Market expansion; Legal, advisory and information assistance; Support for human resource development has really opened the door for small and medium foreign investors.

Incentives when investing in a number of supporting industries according to the policy of developing a number of supporting industries, according to Decree 111/2015/ND-CP dated November 3, 2015, on development of supporting industries.

Thirdly, incentives to invest in agriculture and rural areas according policies to encourage enterprises to invest in agriculture and rural areas: To promote attracting businesses to invest in the agricultural sector, the Government has issued Decree No. 57/2018 / ND-CP (hereinafter referred to as Decree 57) on April 17, 2018 replaces Decree No. 210/2013 / ND-CP with some important financial highlights such as (i) minimizing direct support (in cash), adjusting support levels that are attractive enough to businesses but suitable for the ability to balance the state budget and market mechanism, only partially support total investment in order to create responsibility for effective investment of capital, prevent profit-taking policies; (ii) supplementary support within the authority of the Government as assigned in specialized laws; (iii) the direct support provided in Decree 57 is simpler, smaller in scale and lower in level of support to increase the accessibility of enterprises.

4. Impact of preferential policies to attract foreign direct investment

4.1. Positive effects

In general, the financial policies related to FDI enterprises in Vietnam are largely creating favorable conditions for foreign investors, eliminating discrimination between domestic investors and foreign investor. In which, the Law on Corporate Income Tax and guiding documents applied the law on Corporate Income Tax rates and tax exemption and reduction incentives uniformly for all types of businesses, deregulation of additional corporate income tax and profit remittance tax. The Law on Export Tax and Import Tax has contributed to creating a common and favorable investment environment for both domestic and foreign businesses.

Secondly, the reduction of corporate income tax rates and diversification of tax incentives have contributed to creating a favorable investment environment to attract foreign investment. In the recent 10 years, the amount of foreign direct investment in Vietnam are performing above the level of 10 billion dollars per year. Up to now, Vietnam has attracted investment capital from 130 countries / regions. FDI projects were present in 63/63 localities, FDI capital was also invested in 19/21 business lines of Vietnam.

Thirdly, tax exemptions and reductions have boosted export revenue over the years, especially exports of the FDI sector (Truong and Le, 2016). In particular, a number of FDI projects have provided size large (often projects have been awarded high of the tax incentives) as the project of Samsung Group performed in Bac Ninh and Thai Nguyen province, has made great contributions to the needle Vietnam's export turnover in recent years.

Fourthly, the application of local financial incentives has also brought into play, contributing to positive changes in investment structure by region, especially in the poor central coastal provinces and the Mekong River Delta.

Finally, since 2013, with the passage of the Law of corporate income tax, the way in which incentives are applied has made an important shift from incentives for newly established businesses from investment projects to incentives for income from projects. Investment projects in preferential domains and areas contribute to encouraging enterprises to operate production and business more efficiently, thereby increasing state budget revenues.

4.2. Negative effects

Firstly, corporate income tax incentives are complex and spread. Law of corporate income tax currently defined 3 levels of preferential tax rates based on the industry, incentive sectors and geographical areas of investment and the scale of capital under the provisions of the Investment Law. Specifically, there are 30 sectors to encourage investment and 27 fields to enjoy special incentives when investing, including tax incentives.

To attract investment, the tax incentives in the high being applied to enterprises in economic zones and enterprises to invest in agriculture and rural areas as well as in areas of underdevelopment. In fact, FDI attraction results show that FDI inflows still focus mainly on favorable economic areas, good infrastructure, convenient transportation, seaports... while many areas Economic and industrial zones in underdeveloped areas or like the agricultural sector are highly favored by subjects, but the rate of attraction is very low. As seen, tax incentives are not the most important factor to attract investment. This statement is similar to a study by UNIDO (2011b) that the most important factors that investors are interested in when making investments are in the order of economic and political stability, and labor costs, tax policy, legal framework and quality of infrastructure.

Secondly, Law of corporate income tax incentives in Vietnam are mainly incentives based on corporate profits, in which, incentives for reducing tax rates, applying tax exemption periods, and reducing taxes is the most popular. According to the World Bank (2014), preferential policies on tax exemption and reduction period can encourage businesses to avoid tax through restructuring investments, into new investment projects to continue enjoying incentives. Other forms of tax incentives such as deduction of payable tax amounts and investment-based tax deductions are currently being applied by many countries and have not been applied in Viet Nam.

Thirdly, the provisions on Law of corporate income tax incentives for investment projects are the absolute period from the time the project comes into operation (for tax incentives) and from the time the taxable income is generated from the project investment projects (for incentives and reductions). The tax incentive period does not have any constraints on the duration of the project. This provision easily leads investors to take advantage of preferential policies to benefit, after the incentive period expires, investors can invest in new projects in other localities with the same implementation objectives in order to continue to enjoy endowing.

Fourthly, the situation of dodging the law of import and export tax is quite common. Companies of foreign countries in Vietnam produced to avoid taxes when exporting to other countries under 3 popular formats: make fake certificates of origin (C/O) of Vietnam to enjoy import tax rate lower that the importing country applies to Vietnam compared with the import tax rates applicable to other countries; import complete unit into Vietnam, then pack with the label "Made in Vietnam" and apply for a certificate of origin (C/O) of

Vietnam to enjoy a low tax rate; Investing in a simple factory in Vietnam, then importing almost all foreign accessories and assembling in Vietnam and applying for a certificate of origin (C/O) of Vietnam, although not qualified in terms of value added for export.

Fifthly, the regulations on tax incentives in Vietnam are quite complicated, increasing the cost of tax compliance.

5. Implication

Equations should be numbered consecutively beginning with (1) to the end of the paper, including any appendices. The number should be enclosed in parenthesis and set flush right in the column on the same line as the equation. An extra line of space should be left above and below a displayed equation or formula.

Group of solutions for the overall review of financial policies, especially tax policies to attract foreign direct investment, needs to develop a comprehensive program to review and reassess all. The current preferential policies to attract foreign direct investment capital on both benefits and costs, thereby minimizing the "policy redundancy" causing waste of resources to the state budget. Some specific modifications:

Regarding Law of corporate income tax incentives: policies should be studied in order to switch to a new and effective tax incentive form such as the deduction of tax liability according to the size of capital (disbursement). instead of the scale of investment capital and allowing additional deductions in accordance with the project implementation schedule has been applied by many countries. Focusing on selective incentives, according to subjects such as large projects, national point projects, technology projects, multinational corporations investing in innovation, spillover businesses radiate and increase connectivity with businesses. However, in order not to affect the attractiveness of the investment environment, the application of indirect measures to control the total amount of benefits from Law of corporate income tax incentives to which the investor is entitled.

Import tax policy: Reviewing and re-evaluating the provisions on import and export tax exemption that are prescribed in Article 16 of the Law on Import and Export Tax 2016 towards gradually narrowing the list of commodity tax- free goods groups.

Regarding administrative procedure reform in the tax sector: Continue to reform administrative procedures, reduce tax payment time through the implementation of solutions set out in Resolution No. 02/NQ-CP on improve business environment and enhance national competitiveness. References good experiences in tax administrative procedure reform in the Business Report published annually by the World Bank.

Regarding evaluation of preferential policies: Research and refer to good practices in evaluating investment incentive policies. With regard to tax incentives, Vietnam can apply good standards established by the OECD in the process of reviewing, amending, and proposing new ones as well as implementing and applying tax policies in the coming time.

Group of solutions for anti-transfer pricing and "thin capital"

Research to develop the Law on Transfer Pricing: Vietnam's legal system related to anti-transfer pricing is still very lacking and has many gaps. Currently, the document with the highest legal value related to anti-transfer pricing is Decree 20/2017/ND-CP of the Government regulating a number of contents on tax administration for enterprises with inter-related transactions. Contributing to creating a necessary legal corridor to help the tax authorities to have a basis to control, however, it is necessary to create a clear, healthy and standard tax environment for FDI enterprises, especially for compliance MNCs the formulation of a Law on Transfer Pricing.

To continue to implement solutions to improve the business environment: measures to improve the business environment such as reducing administrative procedures, business conditions; simplify business registration procedures, specialized examination; infrastructure improvements (traffic, information); improving the quality of labor,… are important factors should continue to be made uniform and consistent in this phase.

Group of solutions on policies for sustainable development of FDI capital in the direction of proactively anticipating and selectively attracting the shifting FDI wave in the context of reorganizing the global supply chain. In the short term, it is necessary to: (i) Soon develop and announce policies to attract investment capital flows. In which, it is necessary to clearly state the priority fields and priority localities; (ii) Review of all industrial parks, export processing zones that need to be prioritized for expansion/new construction, industrial parks that need to be narrowed or contracted; publicize the list of industrial parks with clean land fund, ready infrastructure...; and (iii) Strengthening face-to-face meetings between leaders of the Government, Ministries, sectors and localities with major investors to understand the needs and solve problems quickly.

Reference

Aguilera, R. V., & Grøgaard, B. (2019). The dubious role of institutions in international business: A road forward. *Journal of International Business Studies*, 50(1), 20–35.

Bailey, N. (2018). Exploring the relationship between institutional factors and FDI attractiveness: A meta-analytic review. *International Business Review*, 27(1), 139–148.

Donnelly, R., & Manolova, T. S. (2020). Foreign location decisions through an institutional lens: A systematic review and future research agenda. *International Business Review*, 29(4), 101690.

Dunning, J. H. (1977). Trade, location of economic activity and the MNE: a search for an eclectic approach. In B. Ohlin, P. Hesselborn, P. M. Wijkman (Eds.), *The international allocation of economic activity:proceedings of a Nobel Symposium held at Stockholm*, pp.395-418, London: The Macmillan Press Ltd.

Dunning, John (2014), Why Do Companies Invest Overseas? at the page http://www.globalization101.org/why-do-companies-invest-overseas/.

Jain, N. K., Kothari, T., & Kumar, V. (2016). Location choice research: Proposing new agenda. *Management International Review*, 56(3), 303–324.

Jiang, G. F., Holburn, G. L., & Beamish, P. W. (2014). The impact of vicarious experience on foreign location strategy. *Journal of International Management.*

Hans-Rimbert Hemmer, Nguyen thi Phuong Hoa (2002), Contribution of Foreign Direct Investment to Poverty Reduction: The Case of Vietnam in the 1990s, *Univ. Giessen, Fachbereich Wirtschaftswiss.*

Hymer, Stephen H (1960), The International Operations of Foreign Firms: A Study of Direct Foreign Invesment, *Ph.D. thesis*, *Cambridge*, Mass.: MIT Press.

Khachoo, A. Q., Khan, M. I. (2012), Determinants of FDI inflows to developing countries: a panel data analysis, *MPRA Paper 37278, University Library of Munich, Germany*.

Kim, J. U., & Aguilera, R. V. (2015). Foreign location choice: Review and extensions. *International Journal of Management Reviews*.

Kindleberger, C. P. (1969), "American business abroad", *The International Executive*, (11(2), pp.11-12.

Krugman, P.R. (1979), "Increasing Returns, Monopolistic Competition, and International Trade", *Journal of International Economics*, (9(4), pp.469-479.

Meyer, K.E, & Nguyen, H.V (2005). "Foreign investment strategies and sub–national institutions in emerging markets: evidence from Vietnam." *Journal of managenment studies*, 42(1), 63-93.

Nguyen Thi Lien Hoa, Bui Thi Bich Phuong (2014), "Research on factors affecting foreign direct investment in developing countries", *Journal of Development and Integration*, (14 (24), p.40-46.

Nguyen Van Xo (1998), Common Vietnamese Dictionary*, Education Publishing House*.

Nielsen, B. B., Asmussen, C. G., & Weatherall, C. D. (2017). The location choice of foreign direct investments: Empirical evidence and methodological challenges. *Journal of World Business*, 52(1), 62–82.

Solow, R. M. (1956), "A contribution to the theory of economic growth", *The quarterly journal of economics,* (70(1), pp.65-94.

Ministry of Finance (2019), Report on assessment of the current performance of FDI enterprises and current investment incentive policies.

Ministry of Planning and Investment (2018), *Proceedings of 30 years to attract foreign investment in Vietnam*

National Assembly (2020), *Investment Law No. 61/2020 / QH14 dated 17/06/2020*.

National Assembly (2013), Land Law No. 45/2013 / *QH13 dated November 29*, 2013.

Heckscher, E. (1919), "The effect of foreign trade on the distribution of income", *Ekonomisk Tidskriff,* pp.497-512.

Khachoo, A. Q., Khan, M. I. (2012), Determinants of FDI inflows to developing countries: a panel data analysis, *MPRA Paper 37278*, *University Library of Munich, Germany.*

National Assembly (2016), Law on Import and Export Tax No. 107/2016 / *QH13 dated April 6*, 2016.

National Assembly (2013), Law on Corporate Income Tax Amending and Supplementing No. 32/2013 / *QH13 dated June 19*, 2013.

DESIGN AND ANALYSIS OF A MICROGRIPPER BASED ON TRIPLE-STAIR HALF BRIDGE-TYPE COMPLIANT MECHANISM

Thi Diem-My Le [1], Minh Trieu-Nguyen Dang[2], and Chia–Nan Wang [3]

[1]*Department Industrial Engineering Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan R.O.C, ltdiemmy160695@gmail.com*

[2]*Department Industrial Engineering Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan R.O.C, Minhtrieuvietnam@gmail.com*

[3] *Department Industrial Engineering Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan R.O.C, cn.wang@nkust.edu.tw*

Abstract. A new asymmetric microgripper mechanism based on triple-stair half bridge-style (TSBM) is investigated. The design configuration of microgripper is based on a flexure hinge to perform the hold/drop of targets. The corner-filleted flexure hinge is applied in the structure to reduce stress concentration in the corners of the flexure hinges. The output displacement amplification ratio is obtained. The parameters analysis of the asymmetric triple-stair half bridge-style is completed by Finite Element Method in ANSYS software. Then, the Taguchi method is used to optimize the output parameters of the compliance mechanism. The simulation model was carried out to validate the effect of the proposed concept. The optimal result was $l_3 t_1 \alpha_3$ (an incline angle of the flexure hinge of 1.2 degree, a length of flexure hinge of 6 mm, and thickness of flexure hinge of 0.3mm) to obtain the greatest response. The largest output displacement result is 0.5712m, the amplification ratio is 57.

Keywords: Microgripper mechanism, Triple-stair half bridge-type, Taguchi method.

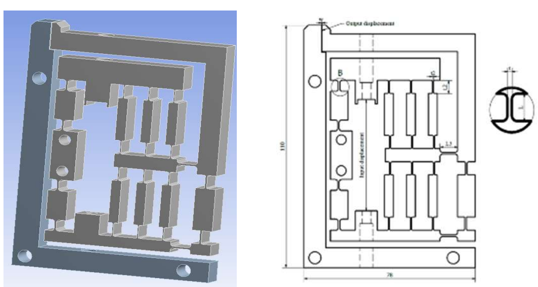
1. Introduction

Microgrippers are developed in micro manufacturing, micro assembling. Zhong Y (2006) proposed the application of microgripper in the field of microsurgery. Maria Chiara Carrozza et al. (2000) presented the micro-gripper applied for manipulating with a typical biomedical micro-device such as the task for the micro-assembly to the accurately position. Nachippan et al. (2018) application model of piezoelectric microgripper for unmanned aircraft was also considered, analysis of the arbitrary variable structure of microgripper was performed and COMSOL MULTIPHYSICS 4.2 software was used for piezoelectric analysis. Research has shown that the material that creates piezoelectric microgripper also significantly affects the degree of movement of the mechanism, specifically compared to conventional materials such as silicon, polysilicon, and silicon dioxide. The result of Silicon dioxide is better than the other materials. The total displacement of the mechanism after structural modification showed a significant increase compared to the existing microgripper. Micro-grippers include two main parts: cantilever and flexible hinge, which ensures that the arms can open and return to their original positions. The first, micro-grippers are designed by using the cantilever. Each arm is fixed at the base, while the material is forced to bend apart or together by a form of actuator (Haddab et al, 2000; Suzuki, 1994). Chen W.et al, (2016) introduced a new hybrid microgripper controlled by PSA and two PCAs. With the proposal of creating a DAM orthogonal table for PSA microgripper, this is a mechanism based on the triangle amplification mechanism with unspecified structural parameters. Microgripper makes it possible to enlarge the gripping stroke and realize the parallel gripping motion in a compact design. The secondly, a microgripper designed with a flexible hinge structure is usually more upscale and has a complex design. They are often developed using methods of mathematical analysis and computer modeling (Zubir et al, 2009), Pseudo Rigid Body Model (PRBM) was also a useful technique to design the geometry of the mechanics and to predict the relative parameter motion to achieve the output responses. Xing Q., Ge Y. (2015), to meet the needs of micro-assembly for micro-tube components, based on the flexure hinge structure a proposed asymmetric flexible microgripper mechanism is proposed. The main structural parameters of the microgripper mechanism are discussed by FEA, the microgripper mechanism is controlled by a piezoelectric actuator, the asymmetric microgripper mechanism operates with the one-sided rule that is fixed without moving, while side holds and moves in parallel using four flexible parallel bars. The microtube will therefore not be destroyed by the homogeneous force caused by the synchronous output displacement of both sides of the microgripper. In this paper, to avoid the disadvantages of the symmetric gripper, the asymmetric structure is proposed to develop a piezoelectric activated microgripper based on a compact flexural hinge with one side of the micro-clamp controlling the displacement while the other side of the micro-clamp is fixed. The asymmetric micro-gripper mechanism could adapt to the micro-operation without damaging.

2. Microgripper structure and simulation

2.1. Design mechanical

The dimension of the specimen is 110x78x10 (mm). Fig. 1a is 3-dimension model, Fig. 1b is 2-dimension drawing. The proposed configuration with the idea is to fix one side of the microgripper jaws, an active party.



a) 3-Dimension b) 2-Dimension

Figure 1. Model design

2.2. Simulation and analysis

The mechanical properties of AL7075 is showed in Table 1. The density is 2810 kg/m^3. The model was designed by using Autodesk Inventor Software and was analyzed by FEA embedded in ANSYS software. The central part of this paper investigates the maximum displacement amplification of the microgripper mechanism.

Table 1: Material of the model

Material	Properties		
	Poisson's ratio	*Young's modulus (GPa)*	*Yield strength (MPa)*
AL7075	0.33	71.7	503

Mesh is created by automatic function and redesigned sizing mesh at the position of FXH to produce the best analysis results as illustrated in Fig.2. And the Fixed support is used for three holes outside and two holes inside. The displacement input controlled by the piezoelectric actuator is shown in Fig.3.

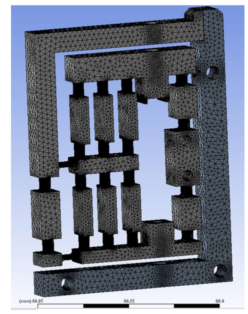


Figure 2. The meshed of microgripper mechanism

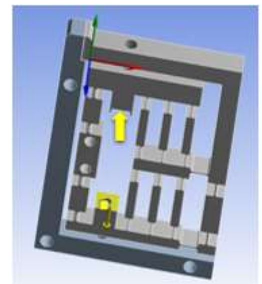


Figure 3. Setup for Displacement

3. Taguchi method

The advantage of the Taguchi method is to reduce the number of tests or simulations by using orthogonal arrays and to minimize the influence of uncontrolled parameters. The loss function is used to find the error between simulated and experimental values. It is called signal-to-noise ratio (S/N) (Taguchi, 1990). It has three concepts: bigger is better, smaller is better and normal is better. In this case, a larger displacement output is better, specifically as follows:

"Greater is better" approach.

$$S/N = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right) \quad (1)$$

4. Results and discussion

4.1. Experiment design

The input parameters of flexure hinge taken for the study include the thickness of flexure hinge, length of flexure hinge, and inclined angle of flexure hinge. The levels of each factor are shown as Table 2:

Table 2: Factor and level

Factor		Unit	Level 1	Level 2	Level 3
Inclined angle of FHs	α	degree	1	1.1	1.2
Length of FH	l	mm	4	5	6
Thickness of FH	t	mm	0.3	0.4	0.5

The Taguchi experimental design involves three factors with three levels, L27 orthogonal arrays were used to conduct the simulation. The observed response is displacement of the output of one side of the microgripper function in the x-axis and the S/N outcome as presented in Table 3, was obtained by equation (1). Table 3 shows the experimental layout and results.

Table 3: Experimental layout using an L27 orthogonal array

No	*α*	*l*	*t*	Result of the simulation	
				Displacement	S/N
1	1	4	0.3	0.58180	-4.70453
2	1	4	0.4	0.45164	-6.90415
3	1	4	0.5	0.35041	-9.10847
4	1	5	0.3	0.58125	-6.35259
5	1	5	0.4	0.45961	-6.75221
6	1	5	0.5	0.3606	-8.85949
7	1	6	0.3	0.57452	-4.81390
8	1	6	0.4	0.46087	-6.72843
9	1	6	0.5	0.36491	-8.75628
10	1.1	4	0.3	0.58232	-4.69677
11	1.1	4	0.4	0.45999	-6.74503
12	1.1	4	0.5	0.36083	-8.85395
13	1.1	5	0.3	0.58149	-4.70915
14	1.1	5	0.4	0.46765	-6.60158
15	1.1	5	0.5	0.37098	-8.61299
16	1.1	6	0.3	0.57465	-4.81193
17	1.1	6	0.4	0.46864	-6.58321
18	1.1	6	0.5	0.37521	-8.51451
19	1.2	4	0.3	0.57899	-4.74658
20	1.2	4	0.4	0.46495	-6.65187
21	1.2	4	0.5	0.3692	-8.65477
22	1.2	5	0.3	0.57787	-4.76340
23	1.2	5	0.4	0.47213	-6.51877
24	1.2	5	0.5	0.37903	-8.42653
25	1.2	6	0.3	0.57121	-4.86408
26	1.2	6	0.4	0.47286	-6.50535
27	1.2	6	0.5	0.3831	-8.33376

4.2. S/N analysis

The average values of S/N respected to levels were recorded in Table 4. In this Table indicated that the maximum values are optimum, namely the optimal S/N value of variable α is -6.607, variable l is -6.657, variable t is -4.940.

Table 4: Response table for signal to noise ratios

Level	*α*	*l*	*t*
1	-6.998	-6.785	-4.940
2	-6.681	-6.844	-6.666
3	-6.607	-6.657	-8.680
Delta	0.391	0.187	3.740
Rank	2	3	1

That result was present in the graph of the S/N plot of displacement as illustrated in Fig. 4, is corresponding with $\alpha_3 l_3 t_1$.

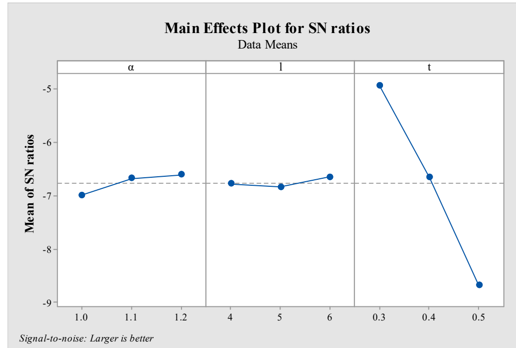


Figure 4. S/N plot of displacement

The mean values of the displacement at the optimal level of design variable as recorded in Table 5, indicated that the maximum values are optimum values of displacement, namely optimal displacement value of variable α is 0.4744, variable l is 0.4718, variable t is 0.5671, is corresponding with $\alpha_3 l_3 t_1$.

Table 5: Response Table for Means

Level	*α*	*l*	*t*
1	0.4540	0.4667	0.5671
2	0.4713	0.4612	0.4643
3	0.4744	0.4718	0.3683
Delta	0.0204	0.0106	0.1989
Rank	2	3	1

The ANOVA analysis shows in Table 6. It's identified that geometry parameters design have significant affection on the frequency. The contributing percent of each parameter are 1.54% α, 0.60% l, 95.37% t and 0.22% deviation error. The P-value makes sure the condition less than 0.05, respectively. This linear regression model fits the data set at level 96.84%.

Table 6: Analysis of variance for transformed response

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
α	2	0.0048	1.54%	0.0048	0.0024	28.17	0
l	2	0.0019	0.60%	0.0019	0.0009	11	0.005
t	2	0.3002	95.37%	0.3002	0.1501	1748.4	0
α*l	4	0.0003	0.09%	0.0003	7E-05	0.82	0.545
α*t	4	0.0034	1.07%	0.0034	0.0008	9.79	0.004
l*t	4	0.0035	1.12%	0.0035	0.0009	10.22	0.003
Error	8	0.0007	0.22%	0.0007	9E-05		
Total	26	0.3148	100.00%				

$$Dis = 0.7393+0.1021\alpha+0.00255l-0.9944t \quad (2)$$

The regression equation (RE) was given out by Minitab 18.0 as Eq. (2). The simulation result and predict result are good correlation as Fig. 5.

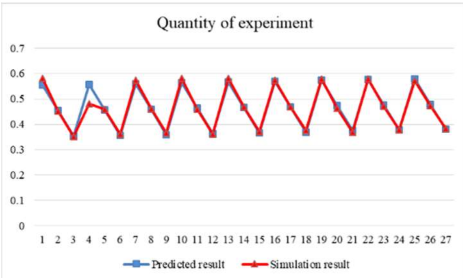


Figure 5. The comparison of the simulation value and predicted value of displacement output.

5. Conclusion

The goal of this study was to achieve the maximum response displacement when optimizing geometric design asymmetric microgripper structure. The L27 orthogonal array was surveyed with three levels of incline angle, length, and thickness of the flexure hinge. The FEA values are fed to the Taguchi method to determine the best parametric geometry to respond to larger displacements. The results are highly satisfied with the design parameters as follows: incline angle (α) is 1.2 degrees, length (l) is 6 mm, thickness (t) is 0.3mm. Displacement is 0.5712mm. In the future, the optimal results are compared with other optimization methods to get the best value. And the optimal results will be verified by measurement data on the actual model.

References

Zhong, Y., et al. (2006). Soft tissue modelling through autowaves for surgery simulation. Med Biol Eng Comput. 44(9): p. 805-21.

Carrozza, M.C., et al. (2000). Towards a force-controlled microgripper for assembling biomedical microdevices. Journal of micromechanics and microengineering. 10(2): p. 271.

Murugu Nachippan, N., A.P. Venkatesh, and M. Muniyappan, (2018). Modelling and Analysis of Piezoelectric Microgripper for Unmanned Aerial Vehicle. Materials Today: Proceedings. 5(9): p. 19456-19462.

Haddab, Y., N. Chaillet, and A. Bourjault. A microgripper using smart piezoelectric actuators. in Proceedings. (2000). IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)(Cat. No. 00CH37113). IEEE.

Suzuki, Y., (1994). Fabrication and evaluation of flexible microgripper. Japanese journal of applied physics. 33(4R): p. 2107.

Chen, W., X. Zhang, and S. Fatikow, (2016). A novel microgripper hybrid driven by a piezoelectric stack actuator and piezoelectric cantilever actuators. Rev Sci Instrum. 87(11): p. 115003.

Zubir, M.N.M., B. Shirinzadeh, and Y. Tian, (2009). Development of a novel flexure-based microgripper for high precision micro-object manipulation. Sensors and Actuators A: Physical, . 150(2): p. 257-266.

Xing, Q. and Y. Ge, (2015). Parametric study of a novel asymmetric micro-gripper mechanism. Journal of Advanced Mechanical Design, Systems, and Manufacturing. 9(5): p. JAMDSM0075-JAMDSM0075.

Taguchi G. (1990). Introduction to quality engineering. Tokyo: Asian Productivity Organization; 1990.

REDESIGNING THE WAREHOUSE'S MATERIAL STORAGE LOCATION: A CASE STUDY OF THE WAREHOUSE OF AN ELECTRIC MOTORCYCLE ASSEMBLY COMPANY

Thi-Huong Pham[1], Quoc-Anh Lai[2], Thi-Thu-Thao Le[3], Thu-Phuong Nguyen[4], Huu-Duc Huynh[5] and Khai-Kien Lao[6]

[1] Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, huong.pham1234567891@hcmut.edu.vn

[2] Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, anh.lai167@hcmut.edu.vn

[3] Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, thao.le11@hcmut.edu.vn

[4] Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, phuong.nguyen20062000@hcmut.edu.vn

[5] Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, hhduc.sdh20@hcmut.edu.vn

[6] Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, khaikien98@hcmut.edu.vn

Abstract. This study focuses on improving the warehouse layout of an electric motorcycle assembly company. In the warehouse area that raw materials and semi-finished products are not properly planned, the moving distance per day is long, and this tends to extend the pick-up time per day. This study combines the 5 Whys method with the Computerized Relationship Layout Planning (CORELAP) algorithm to analyse and rearrange the material placements to minimize the total pick-up distance. The result is that the new design decreases the total moving distance per day by 967.68 meters (from 2682.78 meters to 1715.1 meters). Reducing the total moving distance shortens moving time significantly. This improvement leads to shortening manufacturing time, thus affecting the order fulfillment process of the company positively. Ultimately, the proposed solution is useful for supporting companies in need of redesigning warehouse layout.

Keywords: redesign; warehouse area; total pick-up distance.

1. Introduction

Factors that affect OP (Order Picking) efficiency typically include product demand, warehouse layout, location of items, picking methods combined with routing methods, business employee experience and degree of automation (Gattorna, 1997). The CORELAP algorithm uses the proximity rating stated in the TCR to select the workstation location. This algorithm is a development algorithm (construction algorithm), which is an algorithm used to generate a new layout design that does not depend on or require the original layout. Based on the TCR for each department, where TCR is the number of values calculated based on the ratings of the systems approach (Rahmadani, 2020)

All material handling operations that take place in the warehouse are referred to as warehouse activities. Receiving products, storing them, choosing orders, accumulating and sorting them, and sending them are all examples. In a manufacturing plant, raw materials, semi-finished goods, and final items are stored in manufacturing warehouses. In a warehouse, there is a lot of activity going on.

All warehouse operations are connected to manufacturing since the warehouse is where raw materials are stored. A good warehouse will aid in the quality control of raw materials as well as the production and distribution of items. It will also help to reduce the expenses of production, shipping, and storage.

To effectively manage a warehouse, each warehouse-related activity must be managed, including:

- Arrange and design warehouse premises and storage facilities, as well as warehouse loading and unloading.
- Sorting, finding, classifying, labeling, and disposing of low-quality items are all part of materials management.
- Raw material inventory: correction of inconsistencies (if any), inventory control, and record keeping.
- Import and export of commodities management.
- Ensure mate's safety.

The structure and design of warehouse space is emphasized in the management activities listed above to maximize raw material storage, regulate import and export operations, and material movement material in stock.

We provide an understandable strategy that combines basic algorithms to re-order inventory while simultaneously measuring the model's progress due to the problem's complexity.

This study was carried out at a start-up electric motorcycle manufacturing company in Vietnam. The data collecting procedure lasted for 10 weeks (started in March 2021), so we only focus on organizing the material areas relevant to the assembly process because these areas have a great impact on the time order fulfilment. In our company study, we found some remarkable things: first, the current layout of the company's materials is individually designed for each type of material based on function. Second, when workers pick up materials, they pick them up in groups of parts, each part containing multiple materials; this procedure is time-consuming and sometimes materials are missed. Furthermore, some materials frequently fail during manufacturing and must be regularly picked up, but they are placed far away from the door, so it increases the time and distance of movement. The aim of this study is to support practical problem solving in the company and to lay the foundation for future research and development by redesigning warehouse layout.

Assume that the frequency of import and export is on the order of 60 vehicles per month (based on the average order at the time of data collection) and that the materials required for a one-day assembly are no longer stored at the assembly site.

2. Method

Examine the prospective corporate warehouse's present workflow. To back up the findings, we employed a document review, preliminary interviews, field observations, and historical data collecting. Analyze the root cause of the warehouse problem and offer a remedy. When conducting the investigation. The initial stage is to gather information from a variety of sources, including interviews, historical data, and observations. Following that, metric warehouse workflow mapping is used to assess warehouse process activity and identify linkages that contribute to long lead times. Following the identification of the issue, effective solutions tailored to the issue might be proposed. For the firm to make changes, the proposal would include a thorough warehouse floor layout. The suggestion will be sent to the firm, which will examine it and determine whether or not to implement the solution provided in the research.

2.1. Data collection

Conducting research is determining what kind of data to collect and deciding by what methods that data can be collected. The data collected must be defined and derived from the research goal. When defining data, the following requirements should be observed:

- The information contained in the data must be relevant and enough to clarify the research aim.
- Data must be verified in two ways:
  - Value: the data must quantify the problems that the research needs to quantify.
  - Reliability: that is, repeating the same method must produce the same result.

2.2. The 5 Whys

The 5 Whys approach is an iterative questioning strategy for determining the cause-and-effect link at the root of a problem. The fundamental objective of this method is to find the underlying cause of a flaw or problem by asking "Why?" over and over again. Each response serves as a springboard for the next inquiry. The "5" in the name derives from an observational outcome of the number of iterations required to solve the problem.

The 5 whys method is implemented in 7 steps:

Step 1: Create a Team

Gather together people who are familiar with the details of the problem and with the process, you are trying to fix. Include someone to act as a facilitator who can keep the team focused on identifying effective countermeasures.

Step 2: Identify the problem

If possible, observe the problem in action. Discuss with your group and write a brief, clear statement of the issue on which you all agree. For example, "Team A failed to meet the response time target" or "Software release B resulted in too many rollbacks." Then, write your statement on the board, leaving enough space to write your answer to the repeating question, "Why?"

Step 3: Ask the first question "Why?"

Ask your team why the problem occurred. (Example: "Why didn't team A meet the response time target?")

Step 4: Ask "Why?" Four other times

Working through one of the answers you created in Step 3, ask four more "why" times in a row. Ask a question each time in response to the answer you recorded and record your answer on the right.

Step 5: Know when to stop.

Step 6: Identify the Root Cause

Now that at least one real cause has been identified, it is necessary to discuss and agree on countermeasures that will prevent the problem from recurring.

Step 7: Track the measures.

track how your countermeasures drop or cut the original problem. You may need to change them or replace them with something different. If this happens, it makes sense to repeat the 5 Whys process to ensure that you have identified the correct root cause.

2.3. From-to chart

From - To chart is also known as moving chart or intersection chart between parts on the plan. This chart often shows relationships through numbers, representing the flow of material moving between areas. By examining the data shown on the chart, the designer can determine which areas have a large flow of information or flow between each other when designing, those departments are prioritized for arrangement. or located close to each other. The relationship pairs are all evaluated and assigned to a degree of proximity in the respective scale scales. A used scale is described below: A: Necessary – Necessary, E: Especially important – Especially important., I: Important – Important, O: Should be placed close together – Ordinary closeness, X: Need to be placed far apart – Undesirable.

2.4. Corelap algorithm

Computerized Relationship Layout Planning: A method based on proximity. CORELAP is a construction algorithm based on the conversion of total proximity (TCR) to a score coefficient. A relationship matrix and overall closeness between regions are among the input data. First, calculate the total proximity index (TCR - Total Closeness Rating) to select the part to put in the plan. The TCR for a part is the arithmetic sum of the relationships between that part and all the other parts

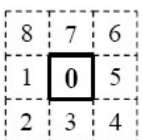
$$TCR = \sum_{j=1, i\neq j}^{m} w_{ij}$$

Next, choose the sort order for the parts. The part that is placed first on the ground is the part with the largest TCR value. If there are multiple choices, choose the part with more A, then E... If there is a component that has an X relationship with the first part, it is arranged last. If 2 or more parts, choose the one with the smallest TCR. The second component is selected as the one with relationship A to the first, if E, I are not considered... If there is more than 1 choice, choose the Part with the largest TCR., if there is a component that has an X relation to a second part, it will be placed next to the last one. If there are many options, choose the one with the smallest TCR. The third part is the one that has the relationship A (if not, consider E, I...) with the arranged parts. If many options are available, select the part with the largest TCR. The algorithm will be repeated until all parts are arranged.

Last step, Layout the parts. Neighborhood: adjacent (having at least one common boundary) or tangent (having a common ground) to the first selected element. Placement rate (PR - placing rating) is the sum of the weighted adjacency ratios between the part to be arranged and its neighbors. With k = {arranged parts}.

$$PR = \sum_{k} w_{ik}$$

The component layout is as follows: the first selected part is placed in the middle. One component layout is determined by evaluating the PR coefficient for all possible locations around the existing premises counterclockwise, starting at the "west corner". The new division is arranged based on the largest PR value.



3. Result and discussion

The case study is the warehouse space of an electric motorcycle assembly company. Total area of warehouse is 104 m2, includes two doors (one door is used for importing and exporting materials, the other one is used for worker moving). The materials placed in the warehouse, are stored base on funtion. Specifically, material storage areas include electronic components, battery, mechanical components, crews, KIT, semi-finished product, defective components, accessary, trash, reseach areas (using rereach and create battery) (Figure 1). The areas are mark on figure 1.

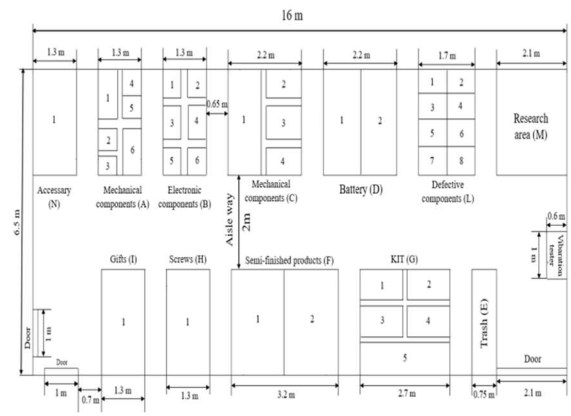


Figure 1: Existing warehouse space

To implement assembly, materials must be moved from warehouse to assembly area. The move is done by a pallet vehical (Figure 2). In one move, the worker will pick up sufficant materials for motorcycle detail with quantity of 10 motorcycles. Details of a motocycle include front wheel, rear wheel, chassis, battery box, turn signal and so on. The company working process as following: at the beginning of workday, the worker will use pallet vehical to take sufficant materials for 1 day (an assemble schedule has been created and worker will use that schedule to determine what the materials will be picked up that day). The picking process is clearly defined, materials will be pick up from the area, which is furthest from the door to the area, which is closest to the door, until there are enough materials. Within the time the worker picks up materials, all of assembly activities at assembly area are almost stopped, it means in the working process have appeared waiting time. This waiting time will affect to fulfill order ability, so the purpose to how to resuce the waiting time in total operting time to improve fufill order ability.

Figure 2: Pallet vehical

3.1. Existing warehouse space assessment

In the company report, average time to pick up materials is approximately twenty minutes per day. This is a lot of time to pick up materials beacause total distance when pick-up proccess is to much. With order quantity of each month is 60 motorcycles. Base on from-to chart (distance between areas) (Figure 3) and order picking for each motorcycle detail, total distance of picking up enough materials to complete 60 motobycles is 2862.7 m (Table 1).

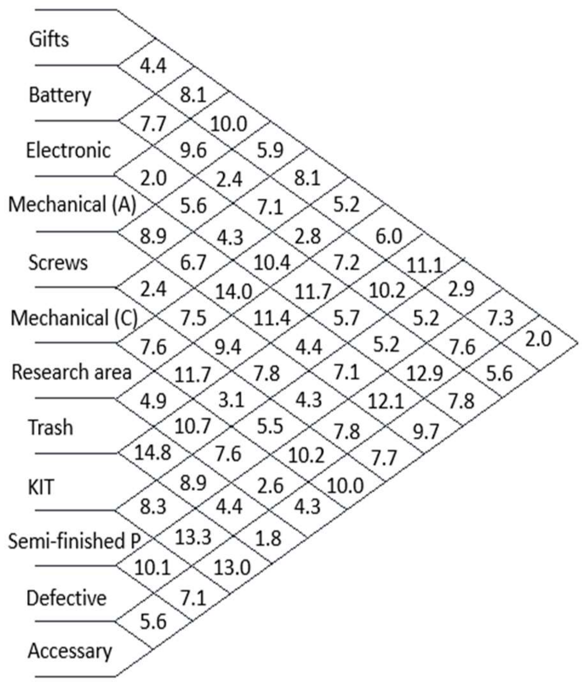


Figure 3: From-to chart

Table 1: Pick-up distance

Car cluster	Detail	Picking process	Distance (m)
Front car cluster	Front wheel	C-G	163.95
	Front fork	C	160.5
	Handlebar assembly	A-H-C-G	271.56
	Front space parts	A	212.7
	Front brake	A-B-H	237.6
	Front turn signal	B-H-C	232.5
Rear car cluster	Rear wheel	C-G	163.95
	Rear brake	A-B-H	237.6
	Rear fork	C	166.8
	Rear turn signal	B-H-C	232.5
	License plates	A	212.7
Battery box	Battery box	B-H-D	229.92
Chassis	Chassis	C	160.5
Total pick-up distance			**2862.7**

The reason of long total distance is areas are not arranged suitably. Based on the picking frequency and the relationship between areas (Table 2), we can see the distances between door and areas, which are visited alot, are further than the distances between door and areas, which are rarely visited and the distances between related areas are quite far.

For example: relationship score=frequency x number of links. Relationship score between A and B area is 12 < B and H area is 30, but distance between A and B area is 1,95m < distance between B and H area.

So, to reduce total pick-up time, we need to reduce total distance. One of the ways is to rearrange the material storage areas in the warehouse.

Table 2: Picking frequency and the relationship between areas

Affiliated areas	Frequency	Number of links	Relationship core	Distance between areas
A-B	6	2	12	1.95
A-C	6	1	6	5.3
A-G	6	1	6	11.03
A-H	6	3	18	5.11
B-D	6	1	6	7.5
B-H	6	5	30	4.25
C-B	6	2	12	2.4
C-G	6	3	18	6.9
C-H	6	3	18	5.3
D-H	6	1	6	7.7
H-G	6	1	6	8.75

3.2. Result of using the algorithm

There are 6 of 12 areas effected to pick-up time includes KIT, mechanical components (A and C), screws and battery so corelap algorithm able for 6 effective areas and the others will be rearrange based on reality warehouse shape.

Using corelap algorithm based on relationship matrix (Table 3) and point assessment table (Table 4) to rearrange the arrangement of areas reasonably to reduce the total pick-up distance.

Table 3: Relationship matrix

Area	A	B	C	D	G	H
A		I	O	X	O	E
B	I		I	O	X	A
C	O	I		X	E	E
D	X	O	X		X	O
G	O	X	E	X		O
H	E	A	E	O	O	

Table 4: Point assessment table

POINT ASSESSMENT TABLE	
Level	Point
A	4
E	3
I	2
O	1
U	0
X	-1

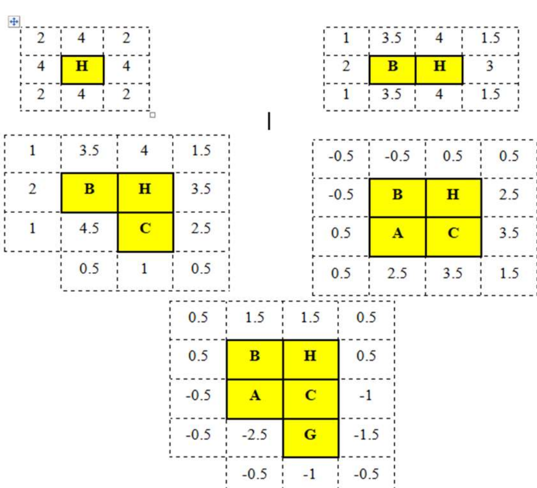
The areas touch each other through 1 point, the adjacent value will be ½ of the original value.

Using CORELAP:

Table 5: Corelap table

	A	B	C	D	G	H	A	E	I	O	U	X	TCR	TT
A		I	O	X	O	E	0	1	1	2	0	1	6	4
B	I		I	O	X	A	1	0	2	1	0	1	8	2
C	O	I		X	E	E	0	2	1	1	0	1	8	3
D	X	O	X		X	O	0	0	0	2	0	3	-1	6
G	O	X	E	X		O	0	1	0	2	0	2	3	5
H	E	A	E	O	O		1	2	0	2	0	0	12	1

Layout order: H – B – C – A – G – D



3.3. Rearrange warehouse space assessment

Based on the results after using the corelap algorithm and the reality warehouse shape, we have after rearrange warehouse space as follows: (Figure 4)

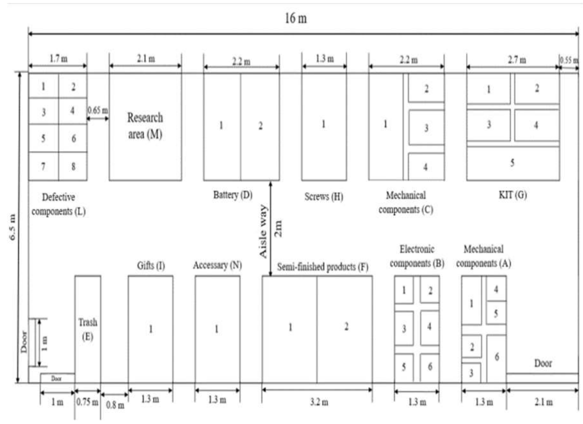


Figure 4: Rearrange warehouse space

After re-arrange the distances between the areas completely changed (we get the new from-to chart) (Figure 5). The material picking process has also been changed to match the company's previous picking method (from the area which is futhest to the door, to the area, which is closest to the door), total distance of picking up enough materials to complete 60 motobycle is 1715.1 m (Table 6). Thus, when the total pick-up distance reduces, the pick-up time reduced, it means we reach the goal of reducing waiting time in total operating time.

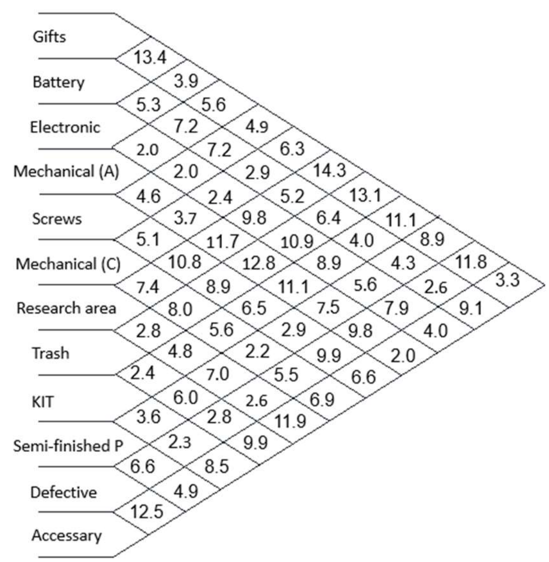


Figure 5: New from-to chart

Table 6: Pick-up distance

Car cluster	Detail	Picking process	Distance (m)
Front car cluster	Front wheel	C-G	111.9
	Front fork	C	111.9
	Handlebar assembly	A-H-C-G	152.7
	Front space parts	A	84.9
	Front brake	A-B-H	158.4
	Front turn signal	B-H-C	164.7
Rear car cluster	Rear wheel	C-G	111.9
	Rear brake	A-B-H	158.4
	Rear fork	C	111.9
	Rear turn signal	B-H-C	164.7
	License plates	A	84.9
Battery box	Battery box	B-H-D	186.9
Chassis	Chassis	C	111.9
Total pick-up distance			**1715.1**

In the future, with the increasing order quantity, the number of materials that need to be picked up every day will also increase. Therefore, the consequences of area rearrangement are waiting time reduce, pick-up time reduce and fulfill order ability enhance.

4. Conclusion

After using Computerized Relationship Layout Planning tool (CORELAP) to improve the warehouse premises. The result is that the total distance to get all the materials in the warehouse after the improvement is B = 1715.1 (m). After improving to complete all orders within 1 month, worker's movement distance has been more significantly reduced than when we compared to the total distance to get all the materials without improvement, which is A= 2771.1 (m).

Material movement time is currently approx 405 minutes cause the worker's velocity is unchanged. Movement distance is reduced resulting in shortened travel time in the warehouse.

The movement distance when loading is reduced, the travel time to get the material is also reduced. Therefore, it decreases the duration of the total production time, meeting the project's goals and the company's requirements.

Reference

Buzacott, J.A. and Shanthikumar, J.G. (1993) *Stochastic Models of Manufacturing Systems*, Prentice-Hall, EnglewoodCliff, NJ.

Chong, T. C., Anderson, D. C., Mitchell, O. R. (1989) QTC - and integrated design/manufacturing/inspection system for poismatic parts. *Proceedings of the ASME Conference on Computers and Engineering, San Francisco, CA,* 417-426.

Frazalle, E.H., Apple, J.M., (1994). "Warehouse Operations", in J.A. Tompkins and D.A. Harmelink (Eds), The Distribution Management Handbook, McGraw-Hill, NY, pp. 22.1-22.36.

Lapedes, A., and Farber, R. F. (1988) *How neural networks work. In D. Z. Anderson (ed), Neural Information Processing Systems (New York: AIP),* chapter 12, 442-456.

Jarvis, J.M., McDowell, E.D., (1991). "Optimal product layout in an order picking warehouse", IIE Transactions, vol. 23, no. 1, pp. 76-87.

Sadeh-Koniecpol, N., Hildum, D., Laliberty, T.J., Smith, S., McA'Nulty, J., and Kjenstad, D. (1996). An integrated process-planning/production-scheduling shell for agile manufacturing. *Technical Report CMU-RI-TR-96-10, Robotics Institute, CarnegieMellonUniversity.*

Swaminathan, J.M., Smith, S., and Sadeh-Koniecpol, N. (1998). Modeling the dynamics of suppy chains: A multi-agent approach*, Decision Sciences,* **29**,607-632.

Sembiring, A. C., et al. (2018). "An application of corelap algoritm to improve the utilization space of the classroom." *Journal of Physics: Conference Series*. Vol. 1007. No. 1. IOP Publishing.

Dwianto, Qodri Azis, dkk. (2016). Usulan Rancangan Tata Letak Fasilitas dengan Menggunakan Metode CORELAP di Perusahaan Konveksi. Jurusan Teknik Industri Itenas No.1, Vol.4.

Rahmadani W. I. (2020). Perancangan Ulang Tata Letak Gudang Menggunakan Metode Konvensional, Corelap dan Simulasi Promodel Jurnal Optimasi Teknik Industri.

Gattorna, J., (1997). Handbook of Logistics and Distribution management, 4th ed., Gower Publisher Company.

Coyle, J.J., Bardi, E.J., Langley, C.J., (1996). "The Management of Business Logistics", 6th ed., WestPublishing, St Paul, MN

ENHANCING THE GARMENT COMPANY'S MANUFACTURING FACILITY LAYOUT

Huu Phuoc Dang[1,*], Phuc Long Nguyen[1], Quoc Cuong Truong[1], Duc Huy Nguyen[1], Huu Duc Huynh[1] and Khai Kien Lao[1]

[1]Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam

**Corresponding: phuoc.dang290120@hcmut.edu.vn*

Abstract. One of the most common problems in garment manufacturers is concentrating on procedures and personnel while neglecting to consider the structure and distribution of workstations within the facility. The proper layout of workstations at a facility aid in reducing travel distance between stations, lowering expenses, and maximizing resource use. In this study, we propose and implement a quantitative research approach in a garment factory utilizing the pairwise position swapping algorithm. We demonstrate that by using the above design solution, we are able to drastically reduce 40% the distance traveled between stations, saving money on transportation, increasing labor productivity, and reducing risk. As a result, the travelling time is decreased accordingly, and the company will be able to increase its competitive position in the domestic production market and, in the future, penetrate the international market.

Keywords: Systematic layout planning; layout design; optimization

1. Introduction

The sensible organization of production stations in a garment manufacturing facility helps to reduce travel distance between stations, lower costs, and maximize resource utilization.

We suggested and executed a quantitative strategy for a garment manufacturing company in this research with the goal of lowering the distance traveled between stations in a single day.

2. Theoretical basis

2.1. Concept of premises and ground dads

The plan is the shape of parts, workstations, equipment, and so on, with a focus on workflow (clients, materials, and so on) across the system.

The process of organizing, arranging, and spatially configuring machinery, equipment, personnel, working areas, and parts for production and service supply is known as layout of production premises in an enterprise.

2.2. Objectives of production premises design

The primary purpose of site design is to make it easier for materials, work, and information to flow through the system. Among the other objectives are the following:

- Create favorable conditions for the production of high-quality goods and services.
- Successful utilization of human resources and workspace.
- Avoid bottlenecks at all costs.
- Cut storage costs for raw materials and semi-finished goods.
- Reduce needless labor and material transfer.
- Reduce production and customer service time.
- Safety is a priority.

2.3. Classification of premises

Product layout, process layout, and fixed site layout are the three primary layout kinds.

2.3.1. Layout by product

Low product unit costs, reduced material transportation distance, reduced labor volume in the process, reduced production time and total production time, streamlined work processes, production system rarely stops due to machinery, equipment, or people failures.

Limitations: Limited flexibility; each time the product is altered, the ground must be adjusted; labor is time and order dependant; monotonous work generates worker boredom; costly machinery and equipment maintenance costs.

2.3.2. Layout of premises according to the process

Process-based layouts group occupations that have comparable processes or functions into departments. In the manufacturing process, the product passes from one portion to the next in the order of the stages that will be done on it.

Arranged according to the process in the form of interrupted production, various types and designs, small volume of each product, orders that change frequently, and the necessity to operate a machine for two or more stages. For example, a hospital is organized by department or professional department, while automotive repair shops organize repair spaces by vehicle part type.

Advantages: high flexibility in equipment and people, small initial equipment investment, improving professional qualifications, diverse jobs that make workers not bored.

Limitations: high unit production costs, inefficient transportation, unstable planning, production schedule, difficulty checking, controlling jobs, low productivity because the jobs are not the same, each time changing workers have to learn new jobs, the level of equipment use is not high.

2.3.3. Layout of premises by fixed position

Positioning is a special layout of a production project in which the product is placed at spot and personnel, machines, and supplies are brought in to execute on-site labor.

This form is ideal for things that are too fragile, big, or heavy to be transferred. For example, when producing aircraft, building ships, doing construction projects, and installing equipment...

Advantages: reduced shipping and shipping expenses, as well as a variety of products and jobs to help alleviate hunger.

Limitations: high expense of shipping equipment, people, and supplies; requires highly skilled and versatile individuals to be able to execute occupations with high levels of specialization.

2.4. Implementation-oriented methods

Based on the analysis of the process and the flow of supplies for the purpose of re-eding the flow of movement within the company premises.

2.4.1. Flow moving in the ground

The flow of moving in the premises can be the flow of materials such as raw materials, sales, unsused products, finished products, documents – procedures, scraps, equipment movement lines, worker movement lines or information. Moving line design principles: bottom-up or small to large principles.

Principle 1: The flow of movement or movement patterns at workstations needs to be designed based on the study of movement and anthropology. Attention should be paid to properties such as:

- Synchronizity: combining the use of hands, arms and legs
- Symmetry: combining motions with body focus
- Naturality: continuous movement, curve and dynamic use
- Rhythmic and habitual: the moving current enables operations to be completed automatically and correctly. It is recommended that you limit your muscle thought, observation, effort, or exertion.

Principle 2: Between workstations, little lines of movement will appear. The interplay between workstations, available space, and the size of the self-lifting device all influence the design of these moving currents.

Principle 3: Plan the flow of movement between functional sections, which is utilized to evaluate the premises' flow of movement or the quality of the premises. It is necessary to take into account the input and output of the system and the flow of movement.

2.4.2. Migration plan

Efficient migration flow planning requires combining the previously described migration lines and selecting reasonable routes to complete the migration process.
To ensure an efficient migration flow, the following objectives must be met:
Increase the number of direct travel routes as much as possible.
Reduce the number of motions being made.
Reduce migration flow expenses as much as possible

2.4.3. Description of the migration line

For premises with only 1 product: Use the migration flow diagram. In this scheme, the steps performed in the order are shown. There are 5 basic steps in processing: processing, moving, checking, waiting and stocking.

For disposal premises at the same time more than one product: Use a multi-product process diagram.

2.4.4. Determination of the flow of migration

From-To charts are frequently used to depict the relationship or movement between elements. The numbers reflecting the flow of materials traveling between machines, workstations, parts, structures, or construction sites are represented in this graph.

This diagram is a type of descriptive model that serves as the foundation for determining premises options. The chart not only indicates the number of moves, but also the distance traveled and the cost of shipping per unit of distance. To determine the exact number of moves, the designer has to know how many units the shipping shipment contains.

2.4.5. Identification of relationships

The designer must consider the factors of a theo magazine in the design of production facilities, such as the tight relationship (i.e., the need to place adjacent parts with each other according to the subjective assessment of the decision-making person).

In a close-up computation, Murther's business relationship chart (REL) is utilized to replace the From-To chart. In the appropriate level scales, all relationship pairs are analyzed and assigned to a neighboring level. The following is a brief description of a common usage scale:

- A: Sym necessary
- E: It's especially important
- I: Quan importanceONedes placed close together
- U: No Matter
- X: Sying far apart

According to the law of experience, the number of X and A relationships should not exceed 5% of the total number of relationships. In turn the maximum cannot be exceeded for scales E, I, O is 10%, 15%, 20%. And occupying the highest level is the U scale when there are at least 50% of relationships out of total relationships.

2.4.6. Building relational diagrams

In the analysis step of the system planning process (SLP), we must create a relationship diagram. The relational scheme's purpose is to define the styledal feature of the components' interaction.

The idea of proximity to the physical aspect to satisfy the relationships between them is the foundation for developing relational diagrams. The importance of the relative position of department pairs is higher, as is the priority level of placement adjacent to each other. On the other hand, the department's relative position is associated with fewer important ties, and the priority level of placing adjacent to one another is likewise lower.

3. Method

Thus, the arrangement and arrangement is not reasonable will cause many losses, . Therefore, it is necessary toorganize theproduction stations in a reasonable and optimal way to achieve the benefits of cost and storage efficiency.

3.1. Ground analysis

3.1.1. Regional data

Table 1: Regional signs

Area symbol	Area name
1	Raw material warehouse
2	Cutting workshop
3	Warehouse drops
4	office
5	iron packing
6	Printing Workshop
7	Garment Factory

3.1.2. Product migration line

3.1.3. Distance matrix

To create a traffic matrix, data on how frequently it flows between stations must be gathered. However, in order to determine the frequency of movement between stations, some information about the output of products generated throughout the manufacturing day of the day is required. Only roughly 882 products are produced at the moment due to the garment factory's capacity, while about 2200 products are outsourced.

Table 2: Traffic information between stations described in the table

Departure area	Arrivals	Traffic of the day
1	2	23
2	6	3
6	7	1
7	5	1
5	3	1

Table 3: Distance between stations

Departure area	Arrivals	Distance (m)
1	2	14
2	6	180
6	7	20
7	5	160
5	3	14

Total distance = distance * traffic

Table 4: Distance matrix

Distance matrix		TO						
		1	2	3	4	5	6	7
FROM	1		14					
	2						180	
	3							
	4							
	5			14				
	6							20
	7					160		

Table 5: Traffic matrix

Distance matrix		TO						
		1	2	3	4	5	6	7
FROM	1		23					
	2						3	
	3							
	4							
	5			1				
	6							1
	7					1		

Table 6: Total distance travel distance between stations of the day

Distance matrix		TO						
		1	2	3	4	5	6	7
FROM	1		322					
	2						540	
	3							
	4							
	5			14				
	6							20
	7					160		

3.1.4. Relationship diagram

Show qua quad (well-founded) relationships through the following levels:

- A - Absolutely necessary
- E - Especially important
- I - Important
- - Should be placed close together
- U - It doesn't matter
- X - Need to be placed apart

Based on the aggregate traffic matrix of all three processes, the team conducts the construction of a graph of relationships. The total travel distance of a product that arrives through each station in a day is categorised by which relationships are based on the following table:

Relationships	Sub-subs	Upper
A	1200	1500
and	900	1200
I	600	900
the	300	600
in the	0	300

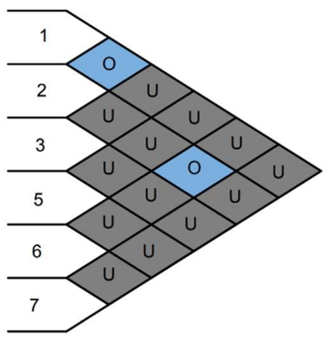


Figure 1: Relationship diagram

3.1.5. Kpanic current way of moving

Currently, to complete a product, from raw materials to packaging, it is necessary to:

Total distance = 322+ 540+ 20+ 160 +14 = 1056 (m)

The travel distance of a product going from the input material to the completion of a product is 1056 (m).

3.2. Ground improvements

3.2.1. Improvement options

Re-allocate stations and use the advanced (improved) algorithm to determine the shortest path.

Improved algorithm by pairing swap method

- Advantages: simple, requirements suitable for current premises
- Cons: this is just a good method that is not the most optimal method

After learning and analyzing, our team chose to renovate the stations in 2 premises and use an improved algorithm because it is suitable for the company's premises.

3.2.2. Redesign of the premises

3.2.2.1. Re-allocation of stations

Currently, the company's premises are vacant quite large so there is no need to care about the area between stations. therefore, the group decided to re-add the ground as follows:

Premises 1: ironing packing, office, warehouse drop, garment factory

Ground 2: printing workshop, cutting workshop, material warehouse.

The total amount of goods moving from station 5 through station 3 is inaligible about (1%) so do not take into account the product movement line

Table 7: Distance between stations

Departure area	Arrivals	Distance (m)
1	2	15
2	6	30
6	7	170
7	5	14
5	3	14

Total travel distance after improvement: 345+90+170+14+14 = 633 (m).

3.2.2.2. Conversion techniques bypairing swapping

First swap: swap positions 6 and 3.

Table 8: Total distance matrix moving between stations in the day after swap

Distance matrix		TO						
		1	2	3	4	5	6	7
FROM	1		345					
	2						90	
	3							
	4							
	5			20				
	6							160
	7					14		

Total distance to travel afterimprovement: 345+90+160+14+20 = 629 (m)

We see that after swapping, the total distance cannot be reduced anymore, so this is the best layout.

4. Conclusions

After observing the current layout and the movement of materials of the product lines at Phuong Nam Garment Trading Co., Ltd., the group noted the actual problems of the company is the relatively long distance of transportation, the ineffective use of the area , not ensuring the safety of workers when working. Therefore, set research objectives as follows:

- Minimize travel distance.
- Optimize the area of use of premises.
- Minimize costs and resources in progress.

After using REL, SLP algorithms. Evaluate the effectiveness of the following criteria:

- Ground usage performance.
- The use of resources.
- The cost of moving materials.
- The level of improvement compared to the current premises.

After improvement, the distance from 1056m to629m, the area is optimally used.

Reference

Buzacott, J.A. and Shanthikumar, J.G. (1993) *Stochastic Models of Manufacturing Systems*, Prentice-Hall, EnglewoodCliff, NJ.

Chong, T. C., Anderson, D. C., Mitchell, O. R. (1989) QTC - and integrated design/manufacturing/inspection system for poismatic parts. *Proceedings of the ASME Conference on Computers and Engineering, San Francisco, CA,* 417-426.

Lapedes, A., and Farber, R. F. (1988) *How neural networks work. In D. Z. Anderson (ed), Neural Information Processing Systems (New York: AIP),* chapter 12, 442-456.

Sadeh-Koniecpol, N., Hildum, D., Laliberty, T.J., Smith, S., McA'Nulty, J., and Kjenstad, D. (1996) An integrated process-planning/production-scheduling shell for agile manufacturing. *Technical Report CMU-RI-TR-96-10, Robotics Institute, CarnegieMellonUniversity.*

Swaminathan, J.M., Smith, S., and Sadeh-Koniecpol, N. (1998) Modeling the dynamics of suppy chains: A multi-agent approach, *Decision Sciences,* **29**,607-632.

Viswajit, T., Teja, T. R., & Deepthi, Y. P. (2017). *A* case study of printing industry plant layout for effective production.

De Carlo, F., Arleo, M. A., Borgia, O., & Tucci, M. (2013). Layout Design for a Low Capacity Manufacturing Line: A Case Study. International Journal of Engineering Business Management.

Ali Naqvi, S. A., Fahad, M., Atir, M., Zubair, M., & Shehzad,M. M. (2016). Productivity improvement of a manufacturing facility using systematic layout planning. Cogent Engineering.

W. Wiyaratn, and A. Watanapa.(2010). Improvement Plant Layout Using Systematic Layout Planning (SLP) for Increased Productivity.

DEMAND PLANNING FRAMEWORK IN VIETNAM FMCG INDUSTRY: A STUDY ON THE MARKETING NEW 4C CONCEPT

Bao Le Gia[1,*] and Thanh-Tuan Dang[2]

[1] *Department of International Business and Economics, University of Greenwich, London SE109LS, United Kingdom*

[2] *Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan*

**Corresponding author: baole7492@gmail.com*

Abstract. Enterprises nowadays focus on effective production and balanced material flows. Under the influence of the 3.0 revolution, the familiar traditional 4Ps model (price, product, promotion, place) has been replaced by the commercialization model of 4Cs, which are customer, cost, convenience, and communication. The inevitable development of industry 4.0 has led to disruptions and challenges that require companies to consider methods of redesigning their supply chain as several emerging technologies are changing the way they do traditional work. The 4Cs model changed and evolved with totally new contents, including co-creation, currency, communal activation, and conversation. Thus, this paper aims to assess the potential factors that influence the current demand planning framework in the Vietnam FMCG industry basing on the new 4Cs concept. The investigation is conducted using a qualitative approach on open-ended questions to converse with respondents and collecting elicit data through focus group and in-depth interview technique. Due to specific research market characteristics such as either the very high price sensitivity of Vietnamese consumers can not wholly reflect changes in demand, buying behavior and decision, or limitations of qualitative method might have resulted in crucial factors contributing to improving the current demand planning framework are unable to answer. The study shows that the new 4Cs marketing concept has a significant impact on the current demand of the Vietnam FMCG industry, especially on the conversion and communal activation. In addition, the demand planning framework in Vietnam has been driven by market characteristics and dependent factors. The study also pointed out some key factors contributing to boost the traditional 3Ps factor: people, process, and product. Finally, this paper has value in ranking and prioritizing the four main elements in marketing 4.0 concerning the demand planning framework in the Vietnam FMCG industry.

Keywords: Marketing 4.0; demand planning; 4Cs concept; Vietnam FMCG.

1. Introduction

The inevitable development of Industry 4.0 has challenges companies to redesign their supply chain as several emerging technologies are changing the way they do traditional work. Advances in information technology, introduction of new products with shorter life cycles, intensified competition in today‘s global markets, and the heightened expectations of customers have contributed to the development of new approaches to Supply Chain as well as Marketing. Above all, the role and power of customers is more important than ever when they are at the center of every business's supply chain. The market, eventually, absorbed the new conception of products and services as enterprises not only bring products to satisfy customers' basic needs and desires, but also pay attention to customers' insight, emotions, and personalization. Then, the product must touch the "heart" of the customer (Rekettye, 2019). New product features allow customers to change their vision of the brand and broaden the relationship between brand image and community, which leading to the three adapted characteristics of Marketing as below (Kotler et al., 2020).

- New concept of branding: instead of advertising brands to sell products, businesses must focus on public relations, environmental commitments and improving reputation through social activities.
- New concept of advertising and communication: providers create advertising content that not only directly talks about brands or products to increase sales, but also creates content that is valuable to customers, serving their personal needs.
- New concept of customer interaction: Moving from the customer service process to the customer care and collaboration process with a wider range of segments, including Netizen.

Along with the development of advanced technology trends, such as e-Commerce, IoT or cloud computing, together with competitiveness in acquiring market share; marketing strategies have been innovated by adapting new changes in targeted customer segment and possible influential factors to customer's buying decision. Apparently, industry experts are required to consider about customer's insight and buying decisions – which demand a significant integration to Marketing 4.0, in order to make the supply chain more responsive and sustainable, instead of focusing on cost optimization or routine collaboration between functional activities like previously (Rekettye, 2019). In most organizations, demand planning plays a strategic role in predicting and planning the product / material requirement according to the business's strategy. This stage requires the cooperation of many departments and under absolute precision in updating and processing data, unifying plans among the boards in the chain. Any inaccuracy will seriously affect the entire supply chain and create an impact on exponential series throughout the operation chain, so known as the 'Bullwhip effect'. The SCOR model from APICS also stressed the importance in managing integrity and optimizing the efficiency of a supply chain, coordination is required from the very first process – Planning (2018).

This paper aims to access the potential factors influencing current demand planning framework in Vietnam – especially in FMCG industry, basing on the new 4Cs concept, by conducting qualitative research method to analyze the three below research questions:

1) *How is the new 4Cs concept influencing market demand?*
2) *How is the current demand planning framework in Vietnam FMCG industry responding to the changing market demand?*
3) *What are the key factors contributing to the improvement of current demand planning framework Vietnam FMCG industry?*

2. Literature review

FMCG industry in Vietnam is a very potential market with rapid growth and recovery (despite the impact of Covid-19), along with trending factors, seasonality, random variation in demand patterns. However, Vietnam is an immature market, with a very high price sensitivity (the Price – Currency factor), promotion-drive (the Promotion – Communication factor), and the demand is easily influenced along the journey of making final purchase decisions. Especially in the context of high standardization, low personalization (the Product – Co-Creation factor) and cost competition, the assessment of the stability of market demand becomes even more complicated. Moreover, Covid-19 also brings consumers closer to new experiences in digitalization (the Place – Communal Activation factor). Shopping online and at mini-supermarkets have reached the highest consumption compared to any previous 4-weeks period. Facebook asserts its position with the highest number of transactions in FMCG trading market share, followed by Shopee – an e-commerce retailer. Both recorded a 3-digit increasing in transaction volume during Covid-19 over the same period (Worldpanel, 2020). Conspicuously, there is a close connection and consensus regarding the shift from Marketing 1.0 to 4.0 within the Demand Planning of Supply Chain Management as illustrated in Table 1. In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis.

Table 1: The evolution marketing 4.0.

Marketing 1.0 and 2.0	Marketing 3.0	Marketing 4.0	Linkage to FMCG product characteristics in Vietnam	Linkage to Demand Planning framework
Product	Customer	Co-Creation	High standardisation, low personalization	Make to Stock, Push system
Price	Cost	Currency	Price sensitivity, Cost competitive advantage	Efficiency supply chain management
Place	Convenience	Communal Activation	Online shopping acceleration, Speed-to-market product	Multi-layer distribution, Retailing pressure
Promotion	Communication	Conversation	Promotion-driven, Immature market	Demand elasticity, ab-normal pattern existed

3. Methodology

According to (Collis, 2014), research is a process of seeking generalized knowledge that can be applied to the explanation of a variety of phenomena. During this process, the source that researcher choose to collect appropriate data and the collecting method is considered as determining factor for the quality of the research analysis. Despite the possibility of data disruption might occur as quantitative method assumes that human behavior and attitudes do not change according to the context (Bryman and Bell, 2011) leading to either lack of involvement of respondents in the issue being surveyed or their perception was framed in the fixed-choice answers; qualitative research method still be identified as an essential one under the research objective due to the following considerations.

- The characteristic of new themes: Qualitative method provides comprehensive information about the characteristics of the social environment in which the study is conducted, and based on a flexible and dialectic research strategy, along with describing and analyzing the behavioral characteristics of groups of people from the researcher's point of view (Collis, 2014). This method allows to detect important topics that researchers may not have covered before which is under agreed perspective with (Saunders et al., 2018) who claimed that one of the four models of situation in qualitative research is inductive thematic – relates to emergence of new codes or themes. As the theory of 4Cs was is still a relatively new concept being explored in terms of its form and applicability to the enterprise (Luciana, 2019), doing a qualitative method in a small and focused scale of samples would prevent the risk of data scattering when the individuals doing the survey have knowledge about this concept, and share their perspective on the same page, thereby leading to more accuracies in data collection.
- The readiness of Vietnam market: Considerations of the context in the country where the research is conducted is also a barrier to the implementation of market survey under quantitative methods. Despite the fact that in recent years "Logistics Management and Supply Chain" and related research topics have been focused by both undergraduate and graduate levels. However, most Vietnamese enterprises (85.7%) have to train and foster logistics manpower by themselves through the on-the-job-training (Vietnam Logistics Report, 2018) because of the shortage of long-term academic education and short-term vocational training. Therefore, in the context of this topic, using quantitative method would hindered by the limitations of academic and market awareness.

Into the bargain, as mentioned above, with a specific research topic requiring knowledge of new Marketing concepts, along with Supply Chain Planning, in-depth interviews with a team of experts were and are working in related topics will help increase the value and authenticity of this research paper (Datt, 2016). Following that, the two techniques of qualitative research method applying for this paper are Focus group and In-depth interview (Semi-structured). The interview questions will be divided into two main parts: part 01 is general questions by which researchers could have better understanding about the background, general recognition of the interviewees regarding to research topic; and part 02 is central part of the interview aiming to overstate the interviewee's professional opinions on the research topic. The set of questions is customized to address one, or two research questions, with a density corresponding to the level of weighted ration, as well as relevance directly to the purpose of the study.

3.1. Focus-group

Focus group interview includes 06 participants, age of 23 to 29, sharing same educational or professional background in Logistics and Supply Chain Management. By accessing the perspectives of those who are formally trained and working properly in professional environment, real-time information from marketplace, along with an acumen in grasping the Marketing trends operating around them - the 4Cs element will be fully exposed. Besides, with solid academic foundation, participants be able to systematize their opinions, as well as reflect the theoretical models mentioned in other academic research. In addition, the diversity of businesses that 06 participants are working with including Third-Party Logistics Services Provider (3PL), F&B - a subsidiary of FMCG, and related academic faculties will provide multi-aspects for the research topic.

3.2. In-depth interview (semi-structured)

Choosing a semi-structured interview format facilitates a clear, seamless conversation that identifies issues to be collected, avoids rambling but still allows the flexibility needed for discussion about new aspects that arise along the line of responses (Clifford et al., 2016). This is a very important factor when selected interviewees of this research are leading experts with career ages spanning from 08 to more than 30. Additionally, their background come from variety of organizations and businesses such as education institutions, e-commerce industry, and FMCG industry; is expected to enrich the quality of data response. However, there might also be a tendency that given information focus primarily on subjective point of view and personal working experience more than objectivity related to other industries, or models, theoretical frameworks.

4. Experiment and results

4.1. Focus group

4.1.1. Agreed perspectives

The focus group makes the following consensus statements on the topic as below.

The pandemic lockdown have shifted shopping behaviors, majority of consumers avoid to shop at crowded places and health care awareness drives changes in customer demand. Consumer prioritizes on the most basic needs, leading to soaring demand for hygiene, cleaning and staples products, while non-essential categories drop-off. Similarly, new shopping pick up action called "Contactless" delivery becomes popular as consumer require delivery man to keep a certain distance a when receiving a product to avoid physical touch.

Marketing plays a vital role to the demand planning by giving planners the marketplace knowledge to align supply and demand considerations using data, information, and analytics. In which, historical data dependency is the biggest constraint. Despite historical data is widely used as a basis for customer behaviors analysis and demand forecast in the rapidly uncertain changing market, it has several limitations that can be addressed as: "historical data can't show market trends and statistical information is not always available on every kind of product or promotion campaign" (interviewee A) and "lead to the lack of information transparency" (interviewee B). Beside the dependency of historical data, "Lack of conversation between sellers and customers and internal functions" is also a constraint that needs to be considered because of "the timely matter in customer demand predictions" (interviewee C) and "the lack of market insights" (interviewee D).

Focus groups all agreed that there is a shifting from the 4P's to the 4C's concept and marketing mix has evolved to accommodate consumer participation in FMCG industry. Both interviewee B and D agree that the transformation can be addressed by higher two-way conversation demand between customer and brand. Conversation is the key ingredient to keep a brand's customers engaged through contextual communication or visual message. Communal activation is a second most important element for this transformation, consumers demand instantaneous access to the products and services they need.

The application of 4Cs concept is agreed to contribute to efficient demand planning by providing quick and timely customer information from conversation, thereby bringing better market and customer insights. For example, "Brand Marketing inputs new trends to support on deciding which SKU to run promotion on" as it would reflect the potential consumption of the item, group of goods, or product and is a two-way flow of information for both the Marketer and the Demand Planner that must be recognized (interviewee D).

The feasibility of implementing the new 4Cs concept in improving the demand planning model can be applied by developing organizational culture that supports conversation between "functions" (interviewee B) and "regional level to local level" as communication in this case plays a critical role (interviewee C).

4.1.2. Different perspectives

Focus-group shows many different views when it comes to assessing which "C factor" will directly affect the market demand through the marketing channel of the business.

Both interviewee B & D agree that Conversation between marketing & demand planning is the most important factor in the 4Cs. Through two-way communication, marketing provides essential marketplace information that impacts changes in demand while planner fulfils the products and services to achieve a high level of customer satisfaction in a rapidly changing business environment. However, while interviewee B focuses on the internal communication throughout an organization will foster collaboration, maintain a customer-centric focus and ensure budgets are properly distributed and support departmental alignment; interviewee D, who working in alcoholic beverages & tobacco area, focus more on proper communication from brand to customer as since 2019, with the promulgation of Decree 100, Vietnamese government officially tightens the use consumption and increase public awareness about the negative effects of alcohol & tobacco. Differently, interviewee A believes that when expand access to international markets, the monetary factor might lead to some unpredictability in the supply chain in the long run. In fact, according to the General Statistics Office of Vietnam, the first 9 months of 2019 witnessed an increase in prices of 10 out of 11 main groups of goods, services, and consumption due to the impact of trade congestion of input materials (Vietnam Credit, 2020). When the dollar rate, the gold price as well as material supply affects the final cost of the product, the Currency better to be set. Another point of view from interviewee C said that Co-Creation should be the most influential factor. By allowing customer personalize products and services, company can increases customer participation in the process of new product development and create superior value propositions. Insight that is generated through this approach will provide more accurate information regarding consumer needs and effectively gain a significant edge over their rivals.

Talking about 2nd place, interviewee A believes that the information exchange between the seller and the buyer will partly support the business better in responding to consumer feedback. Meanwhile, interviewee D thinks that Co-Creation should be the second influential factor as it is essential for delivering a personalized customer experience to each segment of users, and can drive customer loyalty and increase customer satisfaction. Sharing the same perspective, both interviewee B & C put Currency in the 2nd place. interviewee B buying that currency and pricing are not driving just an amount pay for utilizing a product or customers having a service, but also a driving factor in buyer's decision making and seller profitable. Interviewee C pointed out that customers are showing greater price sensitivity in their search for the right product as Covid-19 making consumer priorities become selective on the most basic needs, especially in the scenario of shortage in supplies of goods and materials during the Covid-19.

The next position got majority of agreement by interviewee A, C, and D; that is Communal Activation - a touch point of brand to customer. As E-Commerce expansion has brought the shopping experience to customer fingertips via computers and mobile devices, giving customer easy access to the products and services almost instantly with peer - to - peer distribution and completely changing the way consumers shop. In contrast with others, interviewee B place Co-creation in 3rd position with an argument that customizing and personalizing products and services will enhance customer empowerment in developing new products.

The last position witnessed a clear division when there was no consensus among the four interviewees. Interviewee A, B, and D respectively choose Co-Creation, Communal Activation, and Currency for the fourth place; meanwhile from the perspective of interviewee C, Conversation is ranked last as an element in marketing to convey the brand's message effectively, and thereby contribute to shaping the needs of consumers.

4.2. In-depth interview (semi-structured)

4.2.1. Agreed perspectives

4.2.1.1. The characteristics of Vietnam's FMCG market

Experts agree on FMCG market characteristics in Vietnam as presented in the previous Reports. Especially in terms of the promotion-drive nature, it creates complex forecasting to demand planners. Interviewee 02 stated that promotion is the main factor for consumers to make buying decisions, especially on e-commerce platforms. The competition in terms of service, product quality, or experience is behind the promotion factor when his company has witnessed 10 times higher order growth during each promotion campaign. With that same assumption, interviewee 02 said that the demand characteristic affected by promotion is not only FMCG, but also in all the categories that he has managed, only with different influence levels as the chemical industry does not have a large family name that your company is implementing is mainly B2B. But also, cannot ignore this feature when the products go from inbound going from outbound and approaching the market. Interviewee 03 and 04 gives the view that the word "Immature market" correctly reflects consumer behaviour for FMCG, not only Vietnam in particular, but also as a common characteristic of this industry, when the cost of shifting form buying item X to Y, in the perspective of consumer, is not much worth considering. Interviewee 03 also said that with customers less physically interacting with products to make buying decisions, Marketing's influence through Digital channels could have a far-reaching impact on their "actual" demand and buying decision when all of the touchpoints are now focused on a single click of on customer's smartphone or tablet.

4.2.1.2. Historical data dependence

Dependent on historical data for forecasting is a common judgment among experts when it comes to which demand planning framework is being applied for in Vietnam. The nature of the product, market characteristics, purchasing power, then company strategy and hierarchy are some of the factors that diversify the usage of historical data usage, in which forecasting could be calculated based on 6 months, 12 months or 18 months previously. Interviewee 01 states that there is no perfect pattern, or a proper formula when it comes to demand planning, so the historical data element is just one of the components when running forecasting from the side of planners. And most of all, they need to understand the product and the market they're forecasting - this, none of the data can be accurately reflected, just the judgment and experience of the planner. This statement is on the same page with interviewee 03's response to relying entirely on historical data sometimes as the bottleneck of planners when they use the past to calculate numbers for the future, when consumption trends. An item is in the scope of sales / marketing. Adding to this perspective, interviewee 04 also stated that, as is the case with her business, they confirmed that data of 8 months during the Covid-19 will be removed from all forecasting or even considered as consumer's buying pattern as this is an "anomalous factor". And given the FMCG industry characteristics, when putting an anomaly during forecasting stimulation, the planner's final outcome would turn out to be unreasonable. But as interviewee 02 and 05 argued, this is the common practice of demand planner in Vietnam, and not all businesses clearly recognize the potential of adopting a more optimal planning tool at the moment.

4.2.1.3. Bidirectional communication is indispensable

Bidirectional communication becomes more significant not only between marketing and demand planning functions throughout demand planning process but also between corporate and customer. Interviewee 01 states that data accuracy is the combination of internal information from the supply chain department and external information from sales and marketing, which requires the understanding of common major knowledge and languages through observation and sharing between 2 departments. Meanwhile, both interviewee 02 and 03 believe that by utilizing the marketing department's awareness regarding the target market's needs, interests and challenges, demand planner shall be able to align supply and demand considerations. However, interviewee 04 and 05 emphasized that the deciding factor for bidirectional communication between the Sales / Marketing and the Demand Planner lies in the language used in conversion between the two parties. While Sales / Marketing holds the key factors of demand patterns such as trends, or seasonality which are decision weights in forecasting methods; demand Planner executes the forecasting using technical knowledge and tools based on the historical or on-hand data - this data, sometimes cannot be correctly reflective, or being behind the movement of the market. Therefore, it will take time, patience, and consensus between the two sides to bring the story into the same page to archive the same goal.

4.2.2. Different perspectives

4.2.2.1. Historical Data is not the only constraint

The limitation of running forecasting of the FMCG industry in Vietnam lies not only in the drive-by-promotion characteristics of the market, or heavily depends on historical data, but also the judgment ability of the planner - human capability factor. According to interviewee 01, implementing forecasting for demand planning according to habit and assumption experience is one of the common mistakes of planner. Accordingly, a planner lacks the ability to debate and challenge statistics from sales or market research, so even though his company is currently applying the most advanced systems for Demand Planning (such as DSFM - Demand Sales Forecast management, APODP - Advanced process optimization demand planning) , its limitations are still remaining. Besides, this thinking also leads to resistance when placing a planner in the exchange of new concepts, the new 4Cs is an example, from the Marketing side, as they shape that forecasting models are now sufficient enough for their daily tasks. In the spirit of consensus with the point of view of interviewee 01, Interview 03 and 04 offer another perspective on dependence on Historical Data. They consider using historical data as the basis of all the forecasting methods, and that quantitative will always give more visual results than qualitative, so the problem to be clear is that it's not historical data but how you clean up the historical data before running Forecasting. And communication here is considered a necessary skill than ever to transform the language of Sales and Marketing through the Supply Chain department. According to interviewee 03, she spent a lot of effort in setting up a SOP for the Demand Planning department at the enterprise about Clean Historical data before running Forecasting (Figure 1). The "Cleaning phase" is where Sales and Marketing can openly discuss with Demand Planner about the recorded metrics, along with pointing out factors such as consumer trends, seasonality, about the shift when making buying decisions under the impact of External information should also be provided to Planner. From there they were clearly able to analyse more in-depth the numbers and clean it up to varying degrees, before putting it into the system. She also affirmed that for forecasting, in order to be able to plan for good demand, the simplest method will be the most effective. This opinion is completely agreed by interviewee 05 as he mentioned that the biggest problem of most Forecasting methods is validated data and source of data. Hence, Planner is required to track and clean the data very carefully before inputting to the system because garbage in, garbage out.

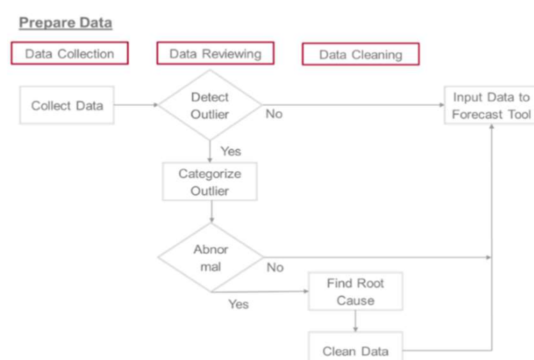


Figure 1. Forecasting process and prepare data.

As for interviewee 02, Historical data is important, but for his industry, it is not a problem because E-commerce has allowed transparency of all purchases and transactions. In addition, purchase metrics, and factors surrounding customer decision-making are clearly represented on a digital platform, and you can fully track them on the flatform. Therefore, the proficiency of data is confirmed to have created a competitive advantage for e-commerce in Vietnam. The problem they are facing is that they still depend on Promotion to generate demand, which has been mentioned as a significant factor affecting the maturity of the market. According to Nielsen digital shopper trend 2018, 85% of customers look for special offers or promotion on special occasions on E-Commerce channels and 73% of customers check and compare with other channels. This dependence makes market characteristics become more and more difficult to forecast. Moreover, sales promotion activities may bring several negative impacts on data accuracy due to the number of competitive promotions. New approaches are promptly cloned by competitors, as each marketer tries to be more creative, more attention getting, or more effective in attracting the attention of consumers and the trade. Therefore, businesses all accept initial losses when investing in e-commerce to capture market share, and the race between e-commerce in Vietnam is like the battle of burning money for businesses (Hong Duong, 2020). Looking at the 03 big players in the Vietnam market, currently, we could claim that Lazada's approach is to develop users and grow quickly, corresponding to accepting losses at the sales stage. Tiki's side chose to build better operations in delivery, product diversification, and customer care and logistics technology in order to achieve customer's loyalty and satisfaction. Shopee, on the other hand, chooses to reduce the delivery cost, product diversification for scaling up. Especially in the context of the Covid-19 translation, with the increasing trend of online consumption, e-commerce floors spend more and more aggressively in their promotion campaign to gain the market share. However, when the demand for online shopping is high, consumers will also have higher demands on quality, types of goods and services. Therefore, the e-commerce floors must also take a very sure step to not be knocked out of this risky race (Huyen Trang, 2020).

4.2.2.2. One "C" is not suit for all industries

- **Agreed on Conversation**

Interviewee 01 and 05 recognize that Conversation is the most important factor in integration and performance improvement of Demand planning. Specifically, interviewee 01 shared that with the industries he worked with, the nature of the promotion has changed a lot and followed by asking himself - a supply chain worker must also understand the promotion message. Sales and Marketing department plans, then prepares appropriately for the resource and capacity of each factory he is in charge of. It is noteworthy that this factor, whether implementing a company in electronics, home appliances (Phillips) or chemicals (Henkel), has the same proportion of value-added during direct conversation with partners and results in a Commercial contract. In addition, when the product is brought downstream and flows into the distribution channels, the impact of "Conversation" will gradually lie in the hands of customers and especially in the current booming social networks context, it requires all the messages will need to be streamlined, clear and precise. Under the same point of view, interviewee 05 also believes that this factor holds a key position in "justifying the demand forecast". If businesses can create a message that communicates well with their customers, demand will be generated naturally and more sustainably than the promotion's one-way push rule as before. Moreover, Promotion in the past places focused on sales and volume growth, but was unable to create interactions with the supply chain under physical, information and monetary flow. Under a perspective of a professor major in Logistics and Supply Chain Management, he states that, with conversation, it should be two-way communication, from which information will be richer and more courageous so that the business will not follow the demand, but attempt to understand the demand, as a result, they can allocate the resources in an optimal way when a promotion message or a campaign is given.

Linking to the case of Biti's - a local brand of Vietnam as a typical illustration for this opinion. First, Biti's made a survey and study on social media that had 87,000 conversations on the topic of "Home coming on Traditional Tet holidays in Vietnam", and analyzed that the story of traveling or returning to family is more interesting and debated by young people. As a domestic brand with a tight budget, with many years of silence, but showing the spirit of "go and experience" of the young, Biti's Hunter has overcome many challenges to listen and speak instead words of many young people: Going is for Returning. As Biti's first step in participating in Tet's media, Biti's want to gain great attention in the community and give a boost to the brand's rise. Moreover, before that, Biti's Hunter sub-brand has been released for a few months but has not been known by many people. Therefore, when building this Tet media campaign, Biti's also aims to raise consumer awareness of the brand (Brand Awareness). Based on a fairly familiar platform during the Tet season, Homing, Biti's chose to turn things around to find new sides of the story. When almost all brands say New Year's return stories, Biti's chose to tell the story to go away. However, it has to be a story that has to do with the brand spirit and convey a meaningful message. And the idea of "Going is for returning" is clearly a new aspect of Homing that does not go against the crowd because it still has the implications of "returning" and has the potential to become an inherent trend in young people, as well as their concerns every Tet. Bitis has succeeded in creating dialogue and understanding the minds of customers. Success is predictable, with Sales growth of 250% compared to Tet 2017, 60% higher than target.

- **Agreed on Communal Activation**

Into a bargain, interviewee 02 and 03 presents a different perspective that Communal Activation should be a most significant one. In the work of Kotler, he stated that "in a connected world, customers demand access to products and services almost instantly, which can only be served with their peers in close proximity. This is the essence of communal activation". interviewee 02 believe that E-commerce is one of the appropriate implementations of this concept. One of the preeminent advantages of E-commerce, he said, from the perspective of a Product owner from Lazada Vietnam, is to completely streamline the downstream distribution network of the supply chain when cutting down intermediate channels allow products directly from seller to buyer. Theoretical speaking, the omission of intermediate channels has eliminated unnecessary touch-points and additional fees for consumer will be reduced. And for the first time, manufacturers can be given the opportunity to grasp the real needs of the market as the journey from them to consumers is shortened and transparent thanks to the technology of E-commerce platforms. Thus, this will be the C factor that strongly affects and, as well as the most profound, on the demand of the market. The remaining problem, as mentioned, is that businesses still use the push demand mechanism through promotion but cannot convert to the pull system, and Logistics cost in Vietnam is an impossible problem which could be solved within a few years. Looking at the reality, two other difficult the issues facing E-commerce businesses today are how to get goods in the fastest way from seller (Tiki solved this problem with their own trading platform and stored one certain quantity at warehouse, but had to trade-off by inventory holding cost as well as slow down the scale-up of the model), and the second is to how to deliver goods in the fastest way to the buyer, he stated. From there, E-commerce sites can increase customer experience and create sustainable value on their buying journey, and can expect a more stable market demand which is not generated through the impact of promotion.

- **Agreed on Co-Creation**

Lastly, in the view of FMCG brands, interviewee 04 stated that Co-creation is by far the impact one. In Vietnam, Nestlé launched its first business in 1912. Over the decades, Nestlé has become familiar to many generations of Vietnamese consumers. To do this, Nestlé always focuses on R&D and testing products that really match not only the needs and tastes of consumers in that country. Nestlé is also the owner of thousands of brands, including many world-famous brands such as Nescafe, Milo, Nestea, Nesvita, Kitkat, Lavie, Cerelac, Nan, ... Among them, the one worth mentioning is Nestcafe, along with Vinacafe and Trung Nguyen are the three big companies that account for more than 80% of the instant coffee market share in Vietnam, and Nestcafe alone is a third of the market share in Asia, Pacific (Vietnambiz, 2019). A noteworthy point is that Nestcafe is also the only foreign brand among the three big players in the playground where consumers have a very special taste and habit of drinking coffee like Vietnam. As a leading coffee producer in the world, it is not difficult for Nescafe to create a "strong" coffee according to Vietnamese taste. The big barrier that Nescafe had to step through at that time was to persuade consumers to believe in it when the perception was that instant coffee could not create a "strong" flavor like handmade Coffee (Thong Chu, 2015). Interview 04 continues her opinion that, after finding the appropriate product to sell, then follow up by the Conversation factor, but only when Co-creation creates a product with enough potential to enter the market. Another good example that can be connected is TH True Milk. As a latecomer and penetrating the Vietnamese dairy market, which is almost completely protected by Vinamilk, until now, after 10 years, TH True Milk's brand position is undeniable. One of the factors leading to Th True Milk's success is the proper investment and R&D team to create immense diversity in their product lines. TH is a business that has enough courage to bring to the market different products that consumers have not even thought of such as sticky-rice Yogurt, jackfruit Yogurt, walnut almond Milk, or even durian Yogurt, and most recently is fresh coconut Yoghurt - all of which are hard to find elsewhere but almost exclusively for Vietnamese palates. According to the interview with Ms. Anh – R&D Director of TH True Milk, the difficulty of the R&D people in the food and beverage industry lies in the rapidly changing minds and tastes of consumers today as, just yesterday, consumers still love green tea matcha, the next day maybe a new drink like black sugar pearl milk tea has usurped. The work of an R&D person requires to understand what consumers really want, and quickly 'catch the trend' to create new products as if you keep selling a classic product, simply milk without creating new excitement, or unable to give people confidence in the health value that the product brings, then you will quickly get lost or being replaced in the market (Minh Chau, 2019).

5. Conclusion

5.1. Potential implications

Under a perspective of implication for Demand Planning, the research is illustrated by the chart of relation between the new 4Cs influences on Demand Planning and its uncontrollable factors (Figure 2). After analyzing the original data from quantitative research, within the limits of the research, we can outline the following 3Ps.

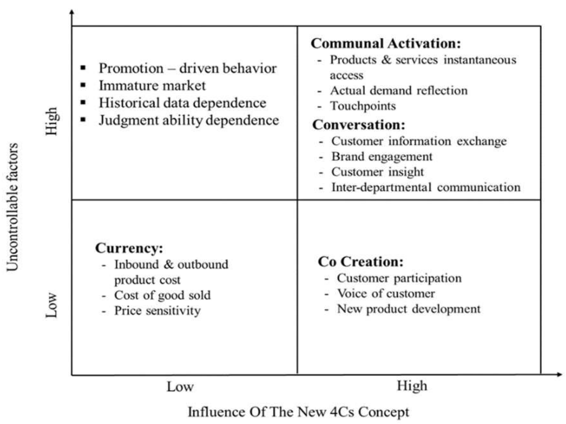


Figure 2. Influences of New 4Cs Concept.

Product: Manufacturing enterprises, or pre-created products in the past are rarely considered as creative leading fields but more about stability, or even conservative. Manufacturing firms strive to achieve stability from quality, productivity, and number of employees to the operational process to reach a threshold called optimization in supply chain operations. However, in this age of industry associated with customer needs, the rapidly changing market has placed even higher demands on businesses, in other words, the increasing requirements of customers, while the cost of switching between products is not significant. Therefore, manufacturing enterprises should not only focus on stabilizing, but also on adapting to change and listening to customers more so that they can create products that not only meet their needs, but also co-creation with consumers.

Process: According to Focus Group's response, and Industry exert 01, 04 and 05 who all mention the importance of communication within the supply chain departments, in this case are Sales / Marketing (external attribution) and Planner (internal attribution). But they also offered a view of how simple communication cannot completely convert information between two fundamentally different functions. Therefore, process, or SOP is considered as a tool that requires it to be applied and monitored closely, not only in the beginning but also throughout the Sales / Marketing process of giving information to Planer, to the technique that Planner offers data, then exchange for adjustment with Sales / Marketing. Interviewee 04 also gave a very realistic case study that in her business, there is one number that Sales / Marketing will need to finalize after forecasting is tolerance - the threshold at which both departments can accept trade-offs which is under their control limit. Thereby she also concluded, agreeing with interviewee 01, the number may be important, but to understand its backend logic, align with the ability to debate and judge the number are prerequisite.

People: People are the factors mentioned throughout the study when talking about the potential applications of the new 4Cs to the Supply Chain. Interviewee 01 and 05 add on the importance of Leadership to change - the factor that should be a pioneer in changing an operating apparatus with many chains is interrelated with each other like the Supply Chain. They emphasize that the application of Conversation, Co-creation, or Communal Activation will not achieve the optimal results if the leader does not recognize its impacts, or resist to innovate, and in-constant along the implementation phase. This also found a consensus from interviewee 03 when she also stated that, although Communal Activation, in the case of Vietnam, it is E-commerce - the playground that no business wants to leave-behind. However, whether the business really shapes its strategy properly and invests in the right people or not is a question to miss. Because of the fact that all three of the big players of the E-commerce race in Vietnam belong to foreign enterprises, they also have the opportunity to hunt for each other's human resources, instead of developing from their core-team. Therefore, the stability of these 03 E-commerce platforms still needs time to give results.
Besides, it is a shortage not to mention the factors that have a great influence on Planners but it is difficult to control. For example, market characteristics and consumers behavior, common practice creates habits about consumer's buying decision, also the capacity to evaluate, process, and debate each planner's data also partly create limits for the application of new knowledge into industry practice.

5.2. Opportunity for further studies

This study is unable to provide fully answer of the 03 research questions due to 02 main limitations below.

- **Specific research market:** as mentioned, Vietnam's consumer market, especially FMCG, an immature market, consumer demand is always influenced by promotion. This characteristic is partly explained by the very high price sensitivity of Vietnamese consumers, so their changes in demand, buying behavior and buying decisions are significantly affected by this weight. Therefore, when placing the research topic in this market, Demand's ideal concepts such as Trend, Seasonality, Cycle will not completely reflect and apply correctly, instead will be Other external drivers, Promotion is in this case.

- **Qualitative research:** in agreement with (Atieno, 2009), when applying qualitative research to a new and in-depth topic, qualitative research will partly help Research data be centralized, selected and unfragmented. However, at the same time the study will not be of great coverage because the method "cannot extend to wider, with the same degree of certainty that quantitative analytics can". Specifically, this research paper is subjective in viewpoints, and experiences of 04 participants for the Focus group, and 05 experts for the In-depth interview. This figure cannot reflect the full amount of information about this topic in the research market.

Thereby, the paper also proposes room for further research within the research question number 03 - the practical potential of the new 4Cs concept in Demand Planning. This information has not been clearly answered, partly due to market characteristics that have formed constraints in the implication and perspective of industry experts. As the in-depth interview also indicates, 4Cs concept has been recognized, personally, in the Supply Chain industry towards their daily work. However, the fact that common practice of the market, under promotion pressure, has been doing well enough to shape consumer demand. Hence, they do not consider the application of this brand-new concept to this immature market promptly. Not to mention the method of communication, leadership, top-down decision in this case will need to be cautiously considered when it might take a certain amount of time to execute a theoretical concept into practical application with result-oriented. Finally, discussions about other aspects in a specific C (Conversation for example), under a relation with inter-communication framework (S&OP for example) to access the level of its impact on consensus forecasting number perhaps are worth studying.

References

Rekettye, G. (2019). Value creation 4.0. marketing products in the 21st century. Transnational press London.

Kotler, M., Cao, T., Wang, S., & Qiao, C. (2020). Marketing strategy in the digital age: Applying Kotler's strategies to digital marketing.

Kantar Worldpanel, K., (2020). Vietnam FMCG insight handbook.

Collis, J., Hussey, R. (2014). Business research: A practical guide for UG and PG students.

Bryman A., Bell, E. (2011). Business research methods Oxford university press.

Saunders, B.,…,& Sim, J., Kingstone. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52, 1893-1907.

Luciana N. Top 5 marketing concepts with examples. Available online: https://planable.io/blog/top-marketing-concepts/ (accessed on 17 January 2021).

Vietnam logistics report. Ministry of industry and trade. Available online: https://vinalinklogistics.com/ban-tin/bao-cao-logistics-viet-nam-2018-logistics-va-thuong-mai-dien-tu-1743.html (accessed on 17 December 2020).

Datt, S. The 8-step procedure to conduct qualitative content analysis in a research. Available online: https://www.projectguru.in/publications/qualitative-content-analysis-research/ (accessed on 18 November 2020).

Clifford, N., Cope, M., Gillespie, T., & French, S. (Eds.). (2016). Key methods in geography. *Sage Journal.*

Vietnam Credit. Available online: https://vietnamcredit.com.vn/vi/san-pham/nganh-nghe-viet-nam/bao-cao-nganh-det-may-48 (accessed on 17 January 2021).

Hong Duong (2020). E-commerce in economic development of Vietnam. Available online: https://tapchitaichinh.vn/tai-chinh-kinh-doanh/thuong-mai-dien-tu-trong-phat-trien-kinh-te-tai-viet-nam-330340.html (accessed on 17 January 2021).

Huyen Trang (2020). The race of e-commerce platforms in Big Sale Month. Available online: https://thegioitiepthi.vn/cuoc-dua-cua-cac-san-thuong-mai-dien-tu-trong-thang-khuyen-mai-khung-182239.html (accessed on 17 January 2021).

Vietnambiz. Available online: https://cdn.vietnambiz.vn/2020/1/20/bao-cao-thi-truong-ca-phe-nam-20191579526264-15795267222611382133113.pdf (accessed on 17 January 2021).

Thong Chu (2015). Nescafe - Vietnamese Cafe and a lesson about understanding. Available online: https://cafebiz.vn/thuong-hieu/nescafe-cafe-viet-va-bai-hoc-ve-su-thau-hieu-20150616111521662.chn (accessed on 17 January 2021).

Minh Chau (2019). Nature Graduated - The core value of TH True Milk. Available online: https://cafef.vn/vong-kim-co-cua-th-true-milk-

20190619151246765.chn (accessed on 19 January 2021).

Atieno, O. (2009). An analysis of the strength and limitations of qualitative and quantitative research paradigms. Masinde Muliro University of Science and Technology, Kenya.

OPPORTUNITIES FOR SUSTAINABLE FASHION: STUDENT SEGMENT IN HO CHI MINH CITY

Le Vu Lan Oanh[1] and Dinh Tien Minh[2]

[1] *Van Lang University, oanh.lvl@vlu.edu.vn*

[2] *University of Economics of Ho Chi Minh city, dinhtienminh@ueh.edu.vn*

Abstract. In recent years, the concept of sustainable fashion is no longer strange in the world. Despite the fact that researchers from all around the world have contributed to this topic, the fashion industry in Vietnam still finds it weird. The topic *"Opportunities for sustainable fashion: student segment in Ho Chi Minh city"* is aimed to determine the level of student awareness about sustainable fashion and to provide sustainable fashion opportunities in the student segment of Ho Chi Minh City. This paper incorporates both qualitative and quantitative research. Two techniques were used to conduct qualitative research: an in-depth interview with two specialists from the University of Architecture and a focus group with students from several universities. A quantitative study was conducted, with empirical data collected from 400 students who study in Ho Chi Minh city. According to the research results, the word "*sustainable fashion*" does not appear to be popular among students in Ho Chi Minh City. In addition, the store-related attributes (SRA), design (DES), and green marketing (GMA) components have all been proven to influence sustainable consumption intention (SCI). For sustainable consumerism, students in Ho Chi Minh City might represent substantial and powerful consumer segment. The research provides insights into the Vietnamese garment business, which may be used to establish strategies for boosting sustainable apparel markets in Ho Chi Minh City and Vietnam as a whole.

Keywords: Sustainable fashion, Green marketing, Sustainable clothing, Sustainability knowledge, Consumer behavior.

1. Introduction

In the midst of a global pandemic as well as horrific natural disasters in Vietnam, the two-year period 2020-2021 ushers in a new but difficult decade. Businesses are being forced to rethink how they did business in the past and consider how to go forward in the future, with fashion being one of the most passionately disputed industries. By 2021, the world's population had surpassed 7.8 billion people (Danso.org, 2021), and resources will be depleted. Fashion can be compared to a gorgeous, sparkling outer shell that shields individuals while also serving as a weapon. However, the brighter something is, the more issues it has.

There have been previous study works on sustainable fashion and the sustainable fashion market, such as *Sustainable clothing: perspectives from US and Chinese young Millennials* (Su, J., Watchravesringkan, K.(T)., Zhou, J. and Gil, M. , 2019), *Sustainable fashion index model and its implication* (Wang, H., Liu, H., Kim, S. J., & Kim, K. H. , 2018), *Shopping for Clothes and Sensitivity to the Suffering of Others: The Role of Compassion and Values in Sustainable Fashion Consumption* (Sonja Maria Geiger & Johannes Keller, 2017) and so on, however, there is a scarcity of study on the intention to buy sustainable fashion.

Objective of this research is: (1) to test and quantify the relationship between sustainable fashion and consumer behavioural intents, (2) to identify opportunities for sustainable fashion among Ho Chi Minh's students and (3) to give the implications for sustainable fashion brands attracting more potential consumers.

2. Literature review

2.1. The concept of sustainability

The Brunt-land report defined sustainability in 1987, and the United Nations' World Commission on Environment and Development (WCED) later endorsed it: *''Sustainability is being able to meet present demands without jeopardizing the ability of future generations to meet their own needs."* (World Commission on Environment and Development, 1987). Sustainability efforts account for the total environmental costs of product manufacture and consumption (Peattie, 2001)Sustainable marketing includes sustainable economic development (Van Dam, Y. K., & Apeldoorn, P. A. C. , 1996) in a sustainable economy.

2.2. Sustainability in the fashion industry

Sustainable fashion is described as apparel that is designed and created with the goal of providing maximum benefits to people and society while minimizing negative environmental consequences (Joergens, 2006). Sustainable fashion is made with biodegradable or recycled materials and environmentally responsible manufacturing procedures, taking into mind its environmental implications (Joergens, 2006); (Fletcher, 2008). Sustainable fashion marketing communication raises consumer knowledge of environmental issues, encouraging individuals to acquire and choose environmentally friendly products, even if they are more expensive (Oliver, J., Benjamin, S., & Leonard, H. , 2019). Consumers' purchase behavior on sustainable fashion is referred to as sustainable fashion consumption (Niinima¨ki, 2010). Recent studies revealed that fashion consumers' sustainable consumption decision is highly complicated (Niinima¨ki, 2010). Despite their favorable attitude toward environmental conservation, fashion buyers are less likely to purchase sustainable apparel (Joergens, 2006); (Niinima¨ki, 2010). There is an attitude-behavior gap of fashion consumers' environmental protection interest and ethical consumption (Niinima¨ki, 2010).

3. Theorical grounding & related research & model framework

3.1. Theorical grounding & related research

The main theory used in this study is the Theory of Planned Behaviour (Ajzen, 1991). Besides, we also use supporting theories such as Balance theory (Heider, 1958) in the article *How to "Nudge" your consumers toward sustainable fashion consumption: An fMRI investigation*. (Lee, E.-J., Choi, H., Han, J., Kim, D. H., Ko, E., & Kim, K. H, 2019); Value-Attitude-Behavior theory (Homer, P.M., Kahle, L.R., 1988) in *Green thinking but thoughtless buying? An empirical extension of the value-attitude behavior hierarchy in sustainable clothing* (Jacobs K, Petersen L, Hörisch J, Battenfeld D, 2018); and Consumer decision-making process theory (Engel, J., Blackwell, R. and Miniard, P., 1986) in *The consumption side of sustainable fashion supply chain* (Chan, T., & Wong, C. W. Y., 2012).

The Theory of Planned Behaviour (TPB) is an extension of the Theory of Reasoned Action (TRA) (Fishbein, M., & Ajzen, I., 1975), (Ajzen, I & Fishbein, M., 1980). Both models are based on the premise that individuals make logical, reasoned decisions to engage in specific behaviours by evaluating the information available to them. The intention of an individual to engage in a behaviour determines the behaviour's execution, the ease with which it can be performed and the views of significant others and the perception that the behaviour is within his/her control. According to the TPB, any action a person takes is guided by three types of considerations: *behavioural beliefs* (beliefs about the probable consequences of the practiced behaviour), *normative beliefs* (beliefs about the normative expectations of other people), and *control beliefs* (beliefs about the presence of factors that may enable or obstruct the performance of the behaviour). Behavioural beliefs usually lead to a positive or negative attitude toward a particular conduct, normative beliefs to perceived social pressure or subjective norms, and control beliefs to perceived behavioural control. The stronger the person's purpose to conduct the activity in question, the higher the favourable behaviour, subjective norm, and perceived control.

According to Balance theory (Heider, 1958) cognitive inconsistency is defined in a different way, with a focus on a triadic relation between the self, another person(s), and an object. Balance theory emphasizes inconsistencies raised by interpersonal relations. Heider's theory is helpful in explaining the complex phenomenon of sustainable fashion consumption. Three general attitudes toward fashion consumption are *sustainability* (entity 1); *product attributes*, such as the *price* and *design* of sustainable fashion products (entity 2); and *individuals' actual purchase behavior* in regard to sustainable fashion (entity 3). Based on balance theory, Lee et colleagues (2019) explain how environmental priming can increase consumer preferences for fashion products with green logos. Using fMRI, they identify the neural representation of the *green logo effect* as significant activations in the anterior cingulate cortex (ACC) (Lee, E et al., 2019).

In Homer and Kahle's the value-attitude-behavior model implies a hierarchy of cognitions in which the influence theoretically flows from more abstract cognitions (i.e., values) to mid-range cognitions (i.e., attitudes) to specific behaviors. It assumes an indirect effect of values on behavior through *attitudes*, i.e. rather abstract values influence more specific attitudes which, in turn, influence particular behavior patterns (Homer, P.M et al., 1988). In contemporary social psychology, an attitude is usually defined as a person's evaluation of a certain object (Fabrigar, L. R., & Wegener, D. T. , 2010). Attitudes can be regarded as key antecedents of behavior and are therefore an integral component of other behavioral models, such as the theory of reasoned action (TRA) (Fishbein, M., & Ajzen, I., 1975) and the theory of planned behavior (TPB) (Ajzen, 1991). Jacobs, K et colleagues used value-attitude-behavior hierarchy as framework, which is augmented by further psychographic constructs hypothesized to influence behavior. Apart from a considerable attitude-behavior gap, the article indicates that a positive attitude towards social-ecological clothing standards, bio spheric and altruistic values, as well as an affinity to online and catalogue shopping, enhance sustainable clothing purchases. Egoistic and hedonic values and, remarkably, a preference for durable clothing hinder sustainable clothing purchase behavior. No significant effects of the suspected barriers – fashion consciousness and price sensitivity – have been identified. The results thus highlight the importance of changing *attitudes and values* towards sustainability, and of focusing on the durability of sustainable clothing and its availability via retail stores (Jacobs K et all., 2018).

Engel, Kollat and Blackwell (1968) describe problem recognition as being caused by a significant difference between a desired state and an actual state, with respect to a particular want or need. While such a conceptualization is useful in understanding problem recognition as the event or "trigger" initiating the consumer decision process, the operational aspects of the precedents to problem recognition, such as the definitions of the 'actual' and 'desired' states and their relative strengths are left unspecified. Similar descriptions of the problem recognition construct are offered by (Wilkie, 1990) and (Hawkins, D. I., Best, R. J., & Coney K. A. , 2001). The goal of this Chan, Y et colleagues' study is to see if there are any links between product and store-related attributes of customers' eco-fashion and consumption decisions, and if such relationships depend on the higher price of eco-fashion. The article indicates that only factors related to the store selling sustainable fashion items positively influence consumers' decision to consume sustainable fashion products, however, the relationship between that system can be undermined by the high prices of sustainable fashion products (Chan, T et al., 2012).

3.2. Model framework

Based on theoretical review, related articles, secondary data and combined with preliminary quantitative survey, the research model is defined as below.

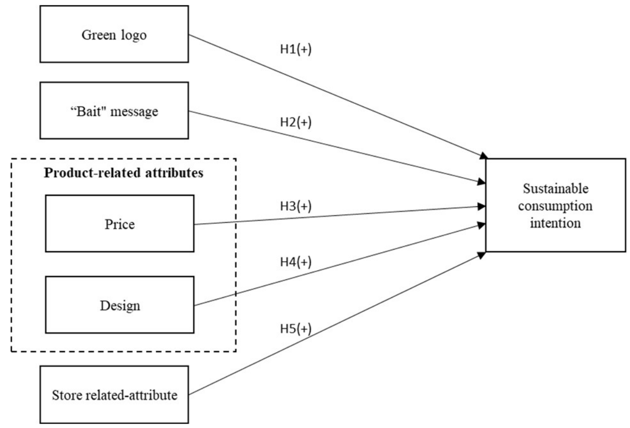


Figure 1: Research model

The first hypothesis is related to the effect of green logos on consumer preference. The most common method green consumers use in making sustainable fashion choices is purchasing products that bear the signs of sustainable quality, such as a GOTS symbol (Global Organic Textile Standard). A green initiative may lead consumers to increase their affinity for a company or a product by generating positive attitudes about that company or products (Minton, A. P., & Rose, R. L., 1997) or Lee et colleagues (2019) and (Yong Jian Wang; Monica D. Hernandez and Michael S. Minor; Jie Wei, 2012) has showed that the impact of green logo on behavior. Therefore, hypothesis 1 is proposed: ***Hypothesis 1: Green logo has a positive impact on the purchase intention of consumers***.

The *"bait" message* can be understood as the meaningful messages that the brand wants to convey to everyone. Business sustainability practices strengthen consumers' perceptions of brands and products, but their impact is not always positive. For example, if consumers assume that a company's sustainability practices are only meant to raise consumer awareness of the environment, their attitude towards the company's sustainability activities will become more negative. more negative (Lois A. Mohr & Deborah J. Webb, 2005); (Sankar Sen & C.B. Bhattacharya, 2001). Many consumers at least philosophically agree with the causes of environmentalism, but when they have to make a realistic decision between an eco-friendly product and a less expensive product, the Their choices are often not inspired by higher causes of sustainability (Carrington, M.J., Neville, B.A. & Whitwell, G.J, 2010). Therefore, implicit messages that attract consumers' attention to information about green and organic products can modify their behavior. The trans-theoretical model of behavior change, the "stages of change" theory in behavioral medicine, suggests that people thinking about adopting lifestyle changes may be motivated to do so. acting with relatively minimal interventions (Marcus, B. H., Selby, V. C., Niaura, R. S., & Rossi, J. S. , 1992). The next hypothesis put forward is: ***H2: The "bait" message has a positive impact on the purchase intention of consumers***.

Product-related attributes of sustainable fashion (PRA) are considered the physical characteristics of a product, related to its features and performance, such as its *design*, *quality*, and *price*. (Keller, 1993). Product-related attributes (PRA) positively influence the purchase of green products (Young, W et al., 2010). Product design, quality and price, play an important role in influencing the intention to buy sustainable fashion products (Butler, S. M., & Francis, S. , 1997); (Solomon, M.R. and Rabolt, N.J., 2004); (Bhaduri, G., & Ha-Brookshire, J. E. , 2011). Fashion consumers won't buy sustainable fashion that they perceive to be overpriced (Lloyd, 1980); (Carrigan, M., & Attalla, A., 2001); (Joergens, 2006). Consumers' low price sensitivity has a positive influence on green purchase intention (Aertsens, J., Mondelaers, K., Verbeke, W., Buysse, J., & Van Huylenbroeck, G., 2011); (Eze, U. C., & Ndubisi, N. O. , 2013); (Lea, E., & Worsley, A., 2008). Therefore, ***H3: The price has a positive impact on the purchase intention of consumers***.

Furthermore, sustainable fashion product design defines the fashion consumer's first impression of the garment and conveys the advantages to the fashion consumer (Bloch, 1995); (Creusen, M. E. H., & Schoormans, J. P. L., 2005). Fashion consumers appreciate stylish clothing with an aesthetically pleasant appearance, therefore sustainable fashion product design symbolizes the value of a fashion department (Joergens, 2006); (Beard, 2008). When fashion consumers are given the option of choosing between two models of fashion clothes with the same price point and functionality, they prefer to choose the more aesthetically pleasing one (Bloch, 1995); (Creusen, M. E. H., & Schoormans, J. P. L., 2005). So, we supposed ***H4: The design has a positive impact on the purchase intention of consumers***.

SRA (Store-related attributes of sustainable fashion) is considered as a characteristic of a store, which is related to store features and store operations (Keller, 1993). Customer service, store design and environment, as well as store ethics and convenience, are all essential variables in influencing purchasing decisions (Erdem, T., Swait, J., Broniarczyk, S., Chakravarti, D., Kapferer, J.-N., Keane, M., … Zettelmeyer, F. , 1999); (Batlas, G. and Papastathopoulou, P., 2003). As fashion consumers perceive a higher quality of customer service, they are more satisfied with the retail store and its products or services, which can then lead to a purchase from the retailer (Cronin, J. J., & Taylor, S. A, 1992); (Zeithaml, V. A., Berry, L. L., & Parasuraman, A., 1996). A welcoming store environment might also encourage customers to make stylish purchases. Ambient (e.g. lighting and music) and social (e.g. employee friendliness) store environment aspects can also impact fashion consumers' shopping decisions (Grewal, D., Baker, J., Levy, M., & Voss, G. B., 2003). Ethical retail activities, such as recycling services and recyclable products, can effectively raise environmental consciousness among fashion buyers (Roberts, J. A., & Bacon, D. R. , 1997). Therefore, fashion consumers consider environmentally responsible SRA as an important criterion when making purchasing decisions (Creyer, 1997); (Carrigan, M., & Attalla, A., 2001). We hypothesize ***H5: The Store related-attributes has a positive impact on the purchase intention of consumers***.

4. Methodology

Leave one-line space after the text and type the table heading, including the table number, above the table. Leave one-line space between the heading and the table. Tables are to be centred on the page. Where possible avoid splitting tables over two pages. A large table or a figure should be positioned at the top of the page (one column, if possible). The study employs a strategy of gathering, analyzing, and synthesizing secondary data from publicly available materials (relevant research articles, websites on linked issues, etc.) to create a theoretical framework. Then, using preliminary quantification, create a formal study model and hypotheses that are relevant to the topic.

The research process was divided into two stages:

- *Preliminary quantitative research:* This stage involves putting the research model and hypotheses to the test with a 100-person online survey. Before doing formal quantitative research, this process determines the correctness, filters factors and variables in the research model, and completes the questionnaire.
- *Formal Quantitative Research:* This is the official research phase to test the research model given earlier. Conduct an indirect interview through online with a sample of 400 people using a simple random sampling approach. Students between the ages from 18 to 24 who live, study, and work in Ho Chi Minh City were interviewed. Finally, data analysis approaches such as descriptive statistics, Cronbach's Alpha, Exploratory Factor Analysis, Pearson correlation and Multiple Regression have made it possible.

Sampling investigation:

The cluster research method (beam) is performed first, which is the preferred method when no list of persons in the population exists. The general population is separated into clusters (beams) whose number of units may be zero or not equal, according to this procedure. People choose a number of clusters to study at random from those clusters, the units are now clusters of units rather than scattered units. Cluster selection was done in two steps: Firstly, six universities in Ho Chi Minh City were chosen, and then individuals were chosen based on clusters (selecting first-year, second-year, third-year and fourth-year students). Then, we performed a stratified sampling method to determine the sample number of each block. Based on the number of enrollments in the last four years of Viet Nam National University Ho Chi Minh City-University of Social Sciences and Humanities (VNU-USSH), Ho Chi Minh City University of Technology (BKU), University of Economics Ho Chi Minh City (UEH), University of Medicine and Pharmacy (YDS), University of Architecture Ho Chi Minh City (UAH) and FPT University (FPT).

Table 1: Number of students in each school

School	Number of students				Total
	first-year	second-year	third-year	fourth-year	
YDS	2708	2520	2192	2062	**9482**
UEH	5500	5000	5000	5000	**20500**
FPT	2200	2200	2000	2000	**8400**
UAH	1500	1500	1200	1200	**5400**
VNU-USSH	3359	3332	3153	2850	**12694**
BKU	5000	5020	5000	4180	**19200**
Total	**20267**	**19572**	**18545**	**17292**	**75676**

Finally, we use the following formula to calculate the sample:

$$n = \frac{N}{1+N(e)^2} \quad (1)$$

n is the sample size
N is the overall quantity
e is the standard error (0.05)

$$n = \frac{75676}{1+7567\ \ (0{,}05)^2} = 398$$

So the sample size equals n = 400. The sample size was distributed among the six universities based on the proportions of first, second, third and fourth year students according to the formula (2):

$$\frac{Number\ of\ students\ \ 〚\text{1st, 2nd, 3rd and 4th year students}\ of\ a\ school〛 \times 400}{Number\ of\ students\ \ 〚\text{1st, 2nd, 3rd and 4th year students}\ of\ the\ six\ schools〛}$$

For example: 1st, 2nd, 3rd and 4th year students of UEH are 20500 people, the number of survey samples at UEH will be: $\frac{20500 \times 400}{75676}$ = 108 (sample)

Table 2: The number of students surveyed per block of each school

School	Number of students				Total
	first-year	second-year	third-year	fourth-year	
YDS	13	13	12	12	**50**
UEH	27	27	27	27	**108**
FPT	11	11	11	11	**44**
UAH	7	8	7	7	**29**
VNU-USSH	17	17	17	16	**67**
BKU	26	26	25	25	**102**
Total	**101**	**102**	**99**	**98**	**400**

5. Result & findings

The study gathered 434 replies to an online survey and chose 400 valid samples, which matched the calculated sample size. In 400 prefiltered samples, a precise calculation was made. There were 149 men and 251 women in all, representing 37.3 percent and 62.7 percent, respectively. Every single one of them is between the ages of 18 and 24, and they all live in Ho Chi Minh City. There are 236 out of 400 people who are aware of sustainable fashion and 164 who are not. Scale testing was performed using Cronbach's Alpha reliability coefficient and exploratory factor analysis (EFA). The results of Cronbach's Alpha analysis (Table 3) show that the scales have high reliability coefficients. The remaining number of observed variables is 23.

Exploratory factor analysis using the Principal Component Analysis method with Varimax rotation to reduce the observed variables into different factors. The result is that the coefficient KMO=0.885 (0.5<KMO<1) of the Barlett test has a significance level of Sig.=000 (<0.05), showing that factor analysis is appropriate. Table 1 presents the components that meet the requirements because the factor loading coefficients of the observed variables are all greater than 0.5 (Hair Jr., J. F. et al. , 1998). The number of observed variables was 23, which were extracted into five (5) factors with Eigenvalues all greater than 1 and the extracted variance reached 61.204%. Therefore, the scales drawn are acceptable, the EFA analysis results also show that the research model has changed, turning the element of the *"bait" message* and *Green logo* into one group, renamed *Green Marketing*. The formal research model includes four (4) independent variables and one (1) dependent variable, and is used for multiple linear regression analysis.

Table 3: Result of Cronbach Alpha and EFA

Variable	Factor				SCI
	SRA	GMA	DES	PRI	
Store-related attributes (SRA) (0.794)					
SRA4	.687				
SRA1	.674				
SRA6	.674				
SRA3	.655				
SRA5	.650				
SRA2	.611				
Green Marketing ("bait" message (0.737) & Green logo (0.737))					
BME1		.763			
BME3		.753			
BME2		.684			
GLO1		.651			
GLO2		.609			
Design (0.822)					
DES3			.768		
DES4			.765		
DES1			.722		
DES2			.708		
Price (0.701)					
PRI1				.695	
PRI4				.693	
PRI3				.675	
PRI2				.670	
PRI5				.640	
Sustainable consumption intention (0.683)					
SCI2					.823
SCI3					.808
SCI1					.712
Eigen Values	5.494	2.290	1.674	1.293	1.836
Cumulative Extracted Variance (%)	27.469	38.917	47.290	53.757	61.204

We tested the hypothesis through multiple regression analysis with SPSS. Table 4 depicts the sig value of PRI > 0.05 (specifically 0.470) so this variable is not significant in the model, sig value of all other variables (SRA), GMA, DES) are all < 0.05 (Table 4). Therefore, the variables SRA, GMA and DES are all significant in the model and have a positive impact on SCI. Thus, the normalized regression equation would be:

F_SCI=0.413*SRA+ 0.326*GMA+ 0.356*DES

In which, SCI: Sustainable consumption intent; SRA: Store-related attributes; GMA: Green marketing; DES: Design.

Table 4: Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1.198E-017	.039		.000	1.000
SRA	.413	.039	.413	10.624	.000
GMA	.326	.039	.326	8.403	.000
DES	.356	.039	.356	9.173	.000
PRI	.028	.039	.028	.723	**.470**

6. Conclusion & recommendation

The study presented an outline of the factors influencing the student segment in Ho Chi Minh city on the sustainable fashion consumption intent. Store-related attributes, in particular, such as friendly store employees, appealing displays, nice colors and sounds in the store, a bustling ambiance, and enough lighting in the store, have the strongest influence with the largest beta coefficient of 0.413, it is one of the most influential elements in determining whether or not a consumer will purchase sustainable clothes. Following is the design factor with a beta coefficient of 0.356, it has shown the importance of style, materials and fabrics in production. Finally, the Green Marketing feature (consisting of two elements: a Green Logo and a "bait" message) with the smallest beta coefficient of 0.326 helps consumers become more aware of the environmental benefits of sustainable fashion items, resulting in increased opportunities for purchase intent. (Table 5).

Table 5: Final result

Factor	Hypotheses	Comments
SRA	SRA is positively associated with SCI	Affected
Green Marketing	GMA is positively associated with SCI	Affected
Price	PRI is positively associated with SCI	NOT affected
Design	DES is positively associated with SCI	Affected

6.1. Level of interest in sustainable fashion among students in Ho Chi Minh City

According to a survey from four hundred (400) respondents, we find out that the number of respondents who have an understanding of sustainable fashion is not high (about 59% of respondents said that they knew about sustainable fashion). Thus, the term sustainable fashion does not seem to be much popular with the student segment in Ho Chi Minh City. With the rapid growth of the Internet and social media, the emergence of international fashion magazine editions in the Vietnamese market, and the fact that fashion design training schools are increasingly developing reasonable, creative curricula, sustainable fashion terminology, and a step-by-step approach to the student segment, first and foremost to individuals with interests and passions in the field, specifically, students from schools with particularities related to fashion such as University of Architecture Ho Chi Minh City, it is easier for them to approach fashion trends, so the number of UAH students who have knowledge about sustainable fashion is quite high (82.8%). Besides, social networking seems to be the most effective way for sustainable fashion to reach closer to the student segment (more than 50% of respondents said that they know the term sustainable fashion through social networking platforms). Despite the fact that there is interest in sustainable fashion, the number of people who have utilized sustainable fashion products is fairly limited (accounting for 15.25 percent). Foreign fashion brands that have entered the domestic market, such as H&M and Uniqlo, have nearly completely dominated the mid-range fashion market share and have gotten a lot of positive feedback from student segment (207 respondents know about sustainable fashion of Uniqlo brand and 187 respondents know about sustainable fashion of H&M brand). Meanwhile, sustainable fashion labels created by Vietnamese designers (Fashion4freedom, Linen, Leinné, Jim Tay, etc.) appear to be unfamiliar to students. However, amid the thousands of various designer fashion labels on the market, these brands continue to survive and develop steadily, indicating that the business philosophy of sustainable fashion is entirely sustainable. It is not a significant barrier to overcome but they need a strategic direction for stable and long-term development in the future.

6.2. Opportunities for Sustainable fashion

Consumer behavior has shifted dramatically as a result of the fast expansion of fashion firms that offer low-cost, stylish clothes. Most items are not used much or are thrown away after only a short time of usage due to damage, ill-fitting, or "out of fashion" according to current shopping patterns. In the global fashion sector, we have created a large amount of garbage, either mistakenly or purposely. Sustainable fashion, in general, is commercial. It can be seen that there are still many opportunities for the fashion industry to develop sustainably in Vietnam in general and the student segment in Ho Chi Minh in particular, some recommendations are given as follows:

For sustainable fashion brands:

- Respond swiftly to a variety of fashion styles, keep up with current trends, and increase client confidence in the product.
- Complicating and globalizing marketing to reach customers.
- Emphasize the importance of store components such as the brand, store layout, sound, and lighting, and consider locating the store in a convenient area for fashion consumers to find.
- Fashion brands can offer services including offering sustainable fashion information, promoting the benefits of sustainable fashion consumption, and boosting consumer confidence. To encourage fashion consumers to make more sustainable purchase decisions, fashion is moving in the direction of sustainable fashion.

For Producers and Distributors:

- Create a complete supply chain like foreign brands
- Synchronous cooperation to bring brands to consumers

Developing online stores, strongly hitting young people's psychology, having to interact continuously and answer questions for customers. At the same time, shopping through an online store can help save time and travel costs.

References

Aertsens, J., Mondelaers, K., Verbeke, W., Buysse, J., & Van Huylenbroeck, G. (2011). The influence of subjective and objective knowledge on attitude, motivations and consumption of organic food. *British Food Journal, 113(11)*, 1353–1378. doi:doi:10.1108/00070

Ajzen, I & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs: Prentice-Hall.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 179-211.

Batlas, G. and Papastathopoulou, P. (2003). Shopper characteristics, product and store choice criteria: a survey in the Greek grocery sector. *International Journal of Retail and Distribution Management, Vol. 31 No. 10,*, 498-507.

Beard, N. D. (2008). The branding of ethical fashion and the consumer: A luxury niche or mass market reality? *Fashion Theory, Vol. 12 No. 4*, 447-468.

Bhaduri, G., & Ha-Brookshire, J. E. . (2011). Do Transparent Business Practices Pay? Exploration of Transparency and Consumer Purchase Intention. *Clothing and Textiles Research Journal, 29(2)*, 135–149. doi:doi:10.1177/0887302x11407910

Bloch, P. H. (1995). Seeking the Ideal Form: Product Design and Consumer Response. *Journal of Marketing, 59(3)*, 16–29. doi:doi:10.1177/002224299505900302

Butler, S. M., & Francis, S. . (1997). The Effects of Environmental Attitudes on Apparel Purchasing Behavior. *Clothing and Textiles Research Journal, 15(2)*, 76–85. doi:doi:10.1177/0887302x9701500202

Carrigan, M., & Attalla, A. (2001). The myth of the ethical consumer--do ethics matter in purchase behaviour? *Journal of Consumer Marketing, 18(7)*, 560–577.

Carrington, M.J., Neville, B.A. & Whitwell, G.J. (2010). Why Ethical Consumers Don't Walk Their Talk: Towards a Framework for Understanding the Gap Between the Ethical Purchase Intentions and Actual Buying Behaviour of Ethically Minded Consumers. *J Bus Ethics, 97*, 139–158. doi:https://doi.org/10.1007/s10551-010-0501-6

Chan, T., & Wong, C. W. Y. (2012). The consumption side of sustainable fashion supply chain. *Journal of Fashion Marketing and Management: An International Journal*, 193–215. doi:doi:10.1108/13612021211222824

Creusen, M. E. H., & Schoormans, J. P. L. (2005). The Different Roles of Product Appearance in Consumer Choice*. *Journal of Product Innovation Management, 22(1)*, 63–81. doi:doi:10.1111/j.0737-6782.2005.00103.x

Creyer, E. (1997). The influence of firm behavior on purchase intention: do consumers really care about business ethics? *Journal of Consumer Marketing, Vol. 14 No. 6*, 421-432. doi:https://doi.org/10.1108/07363769710185999

Cronin, J. J., & Taylor, S. A. (1992). Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing, 56(3)*, 55. doi:doi:10.2307/1252296

Danso.org. (2021, 06 24). *Dan so*. Retrieved from https://danso.org/dan-so-the-gioi/

Engel, J., Blackwell, R. and Miniard, P. (1986). *Consumer Behavior.* New York: The Dryden Press.

Erdem, T., Swait, J., Broniarczyk, S., Chakravarti, D., Kapferer, J.-N., Keane, M., … Zettelmeyer, F. . (1999). Brand Equity, Consumer Learning and Choice. *Marketing Letters, 10(3)*, 301–318. doi:doi:10.1023/a:1008135224357

Eze, U. C., & Ndubisi, N. O. . (2013). Green Buyer Behavior: Evidence from Asia Consumers. *Journal of Asian and African Studies, 48(4)*, 413–426. doi:doi:10.1177/0021909613493602

Fabrigar, L. R., & Wegener, D. T. . (2010). *Attitude structure. In R. F. Baumeister & E. J. Finkel (Eds.),*. New York: Oxford University Press.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Boston: Addison-Wesley.

Fletcher, K. (2008). *Sustainable fashion and textiles: design journeys.* Routledge.

Grewal, D., Baker, J., Levy, M., & Voss, G. B. (2003). The effects of wait expectations and store atmosphere evaluations on patronage intentions in service-intensive retail stores. *Journal of Retailing, 79(4)*, 259–268. doi:doi:10.1016/j.jretai.2003.09.006

Hair Jr., J. F. et al. . (1998). *Multivariate Data Analysis with Readings.* Englewood Cliffs: NJ: Prentice-Hall.

Hawkins, D. I., Best, R. J., & Coney K. A. . (2001). *Consumer behavior: building marketing strategy.* Boston: Irwin/McGraw Hill.

Heider, F. (. (1958). *The psychology of interpersonal relations.* John Wiley & Sons Inc. doi:https://doi.org/10.1037/10628-000

Homer, P.M., Kahle, L.R. (1988). A structural equation test of the value-attitude-behavior hierarchy. *Journal of Personality and Social Psychology, 54*, 638–646.

Jacobs K, Petersen L, Hörisch J, Battenfeld D. (2018). Green thinking but thoughtless buying? An empirical extension of the value-attitude-behaviour hierarchy in sustainable clothing. *Journal of Cleaner Production*. doi:https://doi.org/10.1016/j.jclepro.2018.07.320

Joergens, C. (2006). Ethical fashion: myth or future trend? *Journal of Fashion Marketing and Management, Vol. 10 No. 3*, 360-371.

Keller, K. (1993). Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *Journal of Marketing, 57*, 1-22.

Lea, E., & Worsley, A. (2008). Australian consumers' food-related environmental beliefs and behaviours. *Appetite, 50(2-3)*, 207–214. doi:doi:10.1016/j.appet.2005.07.012

Lee, E.-J., Choi, H., Han, J., Kim, D. H., Ko, E., & Kim, K. H. (2019). How to "Nudge" your consumers toward sustainable fashion consumption: An fMRI investigation. *Journal of Business Research*. doi:doi:10.1016/j.jbusres.2019.09.050

Lloyd, D. (1980). *Demographic factors and mating patterns in Angiosperms. In: Solbrig OT (ed) Demography and Evolution in Plant Populations.* Blackwell: Oxford.

Lois A. Mohr & Deborah J. Webb. (2005). The effects of corporate social responsibility and price on

consumer responses. *Journal of Consumer Affairs, 39(1)*, 121 - 147.
Marcus, B. H., Selby, V. C., Niaura, R. S., & Rossi, J. S. . (1992). Self-Efficacy and the Stages of Exercise Behavior Change. *Research Quarterly for Exercise and Sport, 63(1)*, 60–66. doi:doi:10.1080/02701367.1992.10607557
Minton, A. P., & Rose, R. L. (1997). The Effects of Environmental Concern on Environmentally Friendly Consumer Behavior: An Exploratory Study. *Journal of Business Research, 40(1)*, 37–48. doi:doi:10.1016/s0148-2963(96)00209-3
Niinima¨ki, K. (2010). Eco-clothing, consumer identity and ideology. *Sustainable Development, Vol. 18 No. 3*, 150-162.
Oliver, J., Benjamin, S., & Leonard, H. . (2019). Recycling on vacation: Does pro-environmental behavior change when consumers travel? . *Journal of Global Scholars of Marketing Science, 29(2)*, 266–280. doi:doi:10.1080/21639159.2019.1577158
Peattie, K. (2001). Golden Goose or Wild Goos? The Hunt for the Green Consumer. *Business Strategy and the Environment 10*, 187–199.
Roberts, J. A., & Bacon, D. R. . (1997). Exploring the Subtle Relationships between Environmental Concern and Ecologically Conscious Consumer Behavior. *Journal of Business Research, 40*, 79-89. doi:https://doi.org/10.1016/S0148-2963(96)00280-9
Sankar Sen & C.B. Bhattacharya. (2001). Does Doing Good Always Lead to Doing Better? Consumer Reactions to Corporate Social Responsibility. *Journal of Marketing Research, 38(2)*, 225–243. doi:doi:10.1509/jmkr.38.2.225.18838
Solomon, M.R. and Rabolt, N.J. (2004). *Consumer Behavior: In Fashion.* Upper Saddle River: Prentice Hall.
Sonja Maria Geiger & Johannes Keller. (2017). Shopping for Clothes and Sensitivity to the Suffering of Others: The Role of Compassion and Values in Sustainable Fashion Consumption. *Environment and Behavior*, Fashion Consumption. doi:https://doi.org/10.1177/0013916517732109
Su, J., Watchravesringkan, K.(T)., Zhou, J. and Gil, M. . (2019). Sustainable clothing: perspectives from US and Chinese young Millennials. *International Journal of Retail & Distribution Management, Vol. 47 No. 11*, pp. 1141-1162. doi:https://doi.org/10.1108/I
Van Dam, Y. K., & Apeldoorn, P. A. C. . (1996). Sustainable Marketing. *Journal of Macromarketing, 16(2)*, 45–56. doi:doi:10.1177/027614679601600204
Wang, H., Liu, H., Kim, S. J., & Kim, K. H. . (2018). Sustainable fashion index model and its implication. *Journal of Business Research*, xxx. doi:doi:10.1016/j.jbusres.2017.12.027
Wilkie, W. (1990). *Consumer Behavior.* New York: John Wiley & Sons.
World Commission on Environment and Development. (1987). *Our Common Future.* Oxford: University Press.
Yong Jian Wang; Monica D. Hernandez and Michael S. Minor; Jie Wei. (2012). Superstitious beliefs in consumer evaluation of brand logos : implications for corporate branding strategy. *European journal of marketing : EJM*, 712-732.
Young, W., Hwang, K., McDonald, S. and Oates, C.J. (2010). Sustainable consumption: green consumer behaviour when purchasing products. *Sustainable Development, 18*, 20-31.
Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. *Journal of Marketing, 60(2)*, 31. doi:doi:10.2307/1251929

LOCATING HUBS FOR SAPOTE IN CHAU THANH DISTRICT, TIEN GIANG PROVINCE, VIET NAM

N. V. Chung [1] and H. N. Bao Vi [1]

[1] International University, School of Industrial Engineering & Management, Quarter 6, Linh Trung Ward, Thu Duc City, HCM City, Viet Nam, nvchung@hcmiu.edu.vn

Abstract. The supply chains of tropical fruits in Vietnam currently have many problems leading to reduce the profit for farmers. Specifically, the Mac Bac Sapote in Tien Giang province – Vietnam, it is currently facing challenges such as big fluctuate demand, collecting operation and ineffectively working logistics network and especially the location of Hubs inappropriately. Therefore, this paper emphasized the important of establish and relocation Hub network, it based on the transportation cost, facilities cost of locating, the problem was solved with the main objective is minimizing the total logistics cost and facilities cost in this research.

Keywords: Procurement Hub, Central Warehouse, Location, Logistics Network.

1. Introduction

1.1. Introduction to Mac Bac Sapote in Tien Giang province

Located in the Mekong Delta, Tien Giang is the province which has the largest area and capital of growing tropical fruits in Vietnam. Chau Thanh district, Tien Giang province currently has about 1,874 ha specializing in sapote with an average yield of 25 tons/ha, the output is estimated at 47,043 tons / year. Sapote growing areas are concentrated in communes located along Tien Giang River, south of Chau Thanh district, including communes of Ban Long, Kim Son, Phu Phong, Song Thuan, etc.

With the information from the official website of the People 's Committee of the Tien Giang province. Mac Bac Sapote is a widely known brand in Tien Giang province and Mekong Delta. The fruit is popular and favorable for big fruit, beautiful skin and special sweet. For recent years, the fruit is highly evaluated by not only inland customer but international customer also. However, due to the specifications, the fruit cannot be stored for a long time more than 2 weeks to serve for exporting and there is no solution for extending the storing time of the fruit up to now.

In 2010, the Department of Science and Technology published the report of "Extending and developing Viet Nam sapote", the report pointed out the importance of storing the fruit and growing the fruit with the compliance to Global GAP and Viet GAP. The report also suggested establishing an official organization which can function a warehouse to develop post-harvesting technology to storing the fruit.

1.2. Problem statement

Although Mac Bac Sapote is a special product of Tien Giang province - Vietnam, this kind of fruit is facing with great following difficulties:

- The farmer cannot forecast their customer demand. They just estimating and harvesting as soon as the fruit is ripe. As a result, they will over-supply the product to their customers in the season months and cannot adapt customer demand in the cut-of-season months. Forecasting and producing to adapt customer demand to restrict the fluctuation of the price is what have not implemented yet.
- The problem of the hubs (the places that the farmers sell their fruits), the warehouses are just located without any consideration by the local people. It is the main reason that leads to the big increasing in the transportation cost when the fruit was transported from many places to other places.

This research focusses the demand analysis and forecasting to solve the locating the Mac Bac sapote procurement hub in Chau Thanh, Tien Giang to minimize the transportation cost.

Figure 1. The Mac Bac Sapote

2. Literature review

In one of the initial works in the design of intermodal hub networks, Melkote and Daskin (2001) proposed the integrated model of facility location and transportation network design based on the tradition model of transportation problem and facility location-allocation problem. The model is nonlinear mixed integer programming formulation of incapacitated facility location and network design. The authors assumed that each demands node will be served by the optimal constructed facility without limiting capacity and just considered the location of distribution center and link establishment of transportation.

Meanwhile, Ishfaq and Sox (2011) proposed a mathematical model using the multiple-allocation p-hub median approach. The model encompasses the dynamics of individual modes of transportation through transportation costs, modal connectivity costs, and fixed location costs under service time requirements. There is some limitation in their proposed model such as each load was only allowed visit maximum two hubs in its movement from origin to destination.

In 2015, Sina Rastani et al. (2015) state that hub location problems occur mainly in transportation and telecommunications networks. In this research, the author considered a hub network with capacity constraints on the hubs and hub links. These capacities are multi-level capacities, so hubs and hub links should take a capacity among a set of capacities.

3. Demand forecasting

According to the historical data from the department of Agriculture and Rural Development of Tien Giang province and Mac Bac Sapote Co-operative, the demand and price (sale volume) of the Mac Bac Sapote in the nearest 3 years is shown in the table below.

Figure 2. Demand and Price 2018

Through the graph, we observed that the peak of the demand is often in December May and September while the price seems to be low in those periods. From that, we can conclude that there is a huge problem in supply sources so that cannot satisfy the demand.

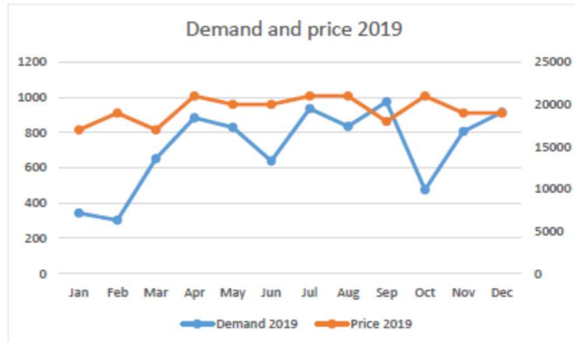


Figure 3. Demand and Price 2019

A forecasting using Winter's model is done to estimate the fruit demand in year 2020. With the level, trend and seasonality of demand, the Winter model is known to be the most accurate forecasting method with the lowest MAD and the most appropriate method S. Chopra and P.Meindl. (2007). Applying $\alpha = 0.05$, $\beta = 0.1$ and $\gamma = 0.1$.

Table 1. Forecasting

Months	Demand 2020	Months	Demand 2020
Jan	803	Jul	460
Feb	724	Aug	489
Mar	687	Sep	601
Apr	916	Oct	905
May	975	Nov	976
Jun	583	Dec	944

4. Hub candidate analysis

4.1. Current supply chain

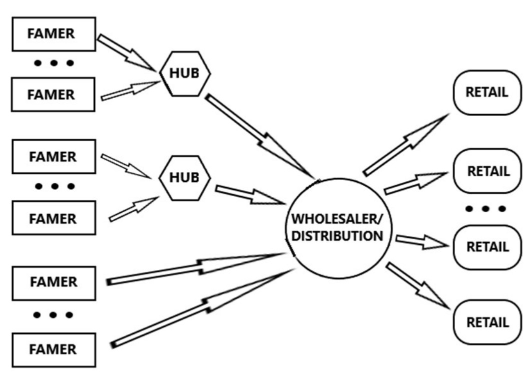


Figure 4. Current supply chain

- Farmer: The main component grows Mac Bac Sapote directly.
- Trader: the owner of hub who collect all Sapote from all regions and organize the transportation to the market.
- Wholesaler/Distribution: Binh Dien Market, Thu Duc Market… the biggest market in Ho Chi Minh City
- Retailer/Customer: Supermarket like BigC, Coopmart, local market, convenience store, etc.

In the current supply chain, there are a waste in transportation from farmer to the market, wholesaler, and distribution due to the transportation scale. Farmer and hub of trader are not connected well, so that individual has to organizer their own transportation. If the farmer and hub of trader are in well connection, it can reduce the transportation cost. The detail about the price is shown in the table below:

Table 2. Pricing of Mac Bac Sapote at each component

	Purchasing price
Farmer	10.000 VND/kg – 20.000 VND/kg
Cooperative	10.000 VND/kg – 20.000 VND/kg
Trader	25.000 VND/kg – 35.000 VND/kg
Wholesale	35.000 VND/kg – 45.000 VND/kg
Distribution	35.000 VND/kg – 45.000 VND/kg
Customer	45.000 VND/kg – 50.000 VND/kg

4.2. Procurement Hub location candidate

Procurement hub is regarded as the "temporary warehouse "and have a function to purchase, collect and concentrate the fruit from local farmers and then transport to the plant. Procurement hub can save cost by improve the cargo efficiency, adjust the transport frequency, and reduce tasks for plant.

- Candidates and Information of proposed procurement hub locations: There are sixteen procurement hub location candidates, which are situated at the communes with good infrastructure of transportation as Vinh Kim, Long Hung, Song Thuan, Kim Son, Dong Hoa, Duong Diem, Phu Phong, Bang Long, Huu Dao, Thanh Phu, Nhi Binh, Tam Hiep, Diem Hy, Binh Trung, Tan Huong, Binh Duc. The other location is not qualified or enough space and infrastructure quality. In these 16 hub locations, there are 12 of them are already exist with small capacity and can be ungraded for increasing capacity. The list of those opened is in the table below.
- Calculating the scale of hub candidate: To determine the area size of hub area in planting area in Tien Giang, it is necessary to apply the land use structure of these centers as the internal land use structure hub services.

According to Japanese Technology Standard Association (www.jsa.or.jp) – the organization that Support of international standardization activities related to ISO, IEC, conduct of Quality Management and Quality Control Certificate Examination. The area of general cargo yard is determined by the following formula:

$$S = N*P/(R*a*W*B)$$

S: Required warehouse - yard area
N: Average demand/year
P: Highest demand coefficient
R: Inventory cycle with year of warehouse
a: area utilization coefficient
B: coefficient of effective storage
W: weight per square area

We have collected of 16 potential locations for hub candidate and apply above formula to calculate the throughput capacity with these parameters:

P = 1.2, R = 12 cycles/year, a = 0.6, W = 2 tons/area, B = 1

Applying the formula with median forecast demand = 9073, we can calculate the total required area around 760 m^2. Compare this value to the real total current operation area is 910 m^2, we can state the real current had enough space for operation.

In table 3, we can see that some candidate hub is already in current supply chain, but we are not sure that they are suitable in our supply chain. So that, in our model will also consider if we should keep or close it for better profit and result.

In addition, we also see some potential locations are not have the facilities yet, and our model also suggest should we build a new hub in that potential location. To create this model as expect, we have collected some operational cost, opening (build) cost, upgrade cost and closing cost.

Table 3. Area of Hub

Candidate Hub	Name	Area total (m^2)	Current using (m^2)	Admin &parking area (m^2)	Current operation area	Through put capacity	Can upgrade/build	Upgraded through put
1	Vinh Kim	200	130	50	80	800	yes	1200
2	Long Hung	220		Not	Build yet		yes	1400
3	Song Thuận	180	120	50	70	700	yes	1100
4	Kim Sơn	120	120	50	70	700	no	700
5	Dong Hoa	150	100	40	60	600	yes	900
6	Duong Diem	180		Not	Build yet		yes	1100
7	Phu Phong	150	150	60	90	900	no	900
8	Bang Long	200	130	50	80	800	yes	1200
9	Huu Dao	130	130	50	80	800	no	800
10	Thanh Phu	150	150	60	90	900	no	900
11	Nhi Binh	150		Not	Build yet		yes	900
12	Tam Hiep	180	130	50	80	800	yes	1100
13	Diem Hy	180	100	40	60	600	yes	1100
14	Binh Trung	200		Not	Build yet		yes	1200
15	Tan Huong	200	130	50	80	800	yes	1200
16	Binh Duc	120	120	50	70	700	no	700
Total current operation area 910 m^2								

In this paper, we choose the best central warehouse/distribution candidate to build a complete Hub network for Mac Bac Sapote. Our criteria for the central warehouse/distribution are in good conditioning, near the national road 1A or near the Trung Luong Highway. After search and survey the take holder, we have three best candidates:

Table 4. Central warehouse/distribution candidate

Warehouse	1	2	3
Throughput capacity (tons)	7000	5000	4000
Opening cost (Million VND)	1,379	1,045	812
Yearly operational Cost (Million VND)	6,790	4,850	3,880

Table 5. Transportation cost (source: survey)

Segment	Cost	Available mode
Farmer - Collecting Hub	21,000 VND/tons/Km	Road, River
Collecting Hub - Central warehouse	19,000 VND/tons/Km	Road
Central Warehouse - Customer	15,000 VND/tons/Km	Road

4.3. Results

The main idea of this model from Wesolowsky, G. O. and Truscott W. G. (1976) and Sina Rastani, Mostafa Setak & Hossein Karimi (2015). After running the model, we get the result about the structure of our system. First is the location and the scale of the hub. As we know, we have total 16 potential locations of hub, 12 over 16 locations have been opened for the collection hub and 5 in 12 hubs have reach the maximum throughput capacity. Due to the increasing in demand, we decide to build the model consider the current situation and run to find the optimal solution. One of that result is about the collecting hub which is shown in the table 7:

Current:

- Total used: 12/16 hubs: Song Thuan, Kim Son, Dong Hoa, Duong Diem, Phu Phong, Thanh Phu, Tam Hiep, Diem Hy, Tan Huong, Binh Duc.
- Reach max capacity: 5/12 used hub.
- Can upgrade/open new: 11/16 hubs.

Solution:

- Total used: 10/16 hubs.
- Reach max capacity: 8/9 used hub.
- Close: 3 hubs.

From the solution, we can also have the solution of central warehouse, we rent 2/3 potential warehouse.

Table 6. Result of central warehouse (VND)

Warehouse	1	2	3
Throughput capacity (tons)	7000	5000	4000
Opening cost (Million VND)	1,379	1,045	812
Yearly operational Cost (Million VND)	6,790	4,850	3,880
Rent result	Yes	No	Yes

In conclusion, we have proposed the new model which can make decision based on the current situation of the supply chain.

5. Conclusion

In this research, the current problem of locating hub for Mac Bac Sapote product in Tien Giang province has been solved. Based on the data input about transportation cost, facilities cost of locating and relocating, the problem was solved by CPLEX with the main objective is minimizing the total logistics cost and facilities cost. In addition, the research also emphasized the important of establish and relocation hub network for better collection task and transportation task. The result showed that some current hub should be closed or upgraded. Some potential location which does not build hub yet, our model also proposed where to build the new one. The different in closing cost, upgrade cost, current operation cost, new upgraded operation cost are also considered in our model. Our solution is applied to solve the real problem for Mac Bac Sapote.

Table 7. Result of location collecting hub.

Hub	Name	Upgraded/Open status	Current capacity	Solution/Decision from model	New capacity	Max capacity
1	Vinh Kim	Can Upgrade	800	Close Hub		
2	Long Hung	Can open	Not open yet	No open		
3	Song Thuan	Can Upgrade	700	Upgrade Hub	1100	Yes
4	Kim Son	No	700	Keep	700	Yes
5	Dong Hoa	Can Upgrade	600	Keep	600	
6	Duong Diem	Can open	Not open yet	Open Hub	1100	Yes
7	Phu Phong	No	900	Keep	900	Yes
8	Bang Long	Can Upgrade	800	Close Hub		
9	Huu Dao	No	800	Close Hub		
10	Thanh Phu	No	900	Keep	900	Yes
11	Nhi Binh	Can open	Not open yet	No open		
12	Tam Hiep	Can Upgrade	800	Keep	800	
13	Diem Hy	Can Upgrade	600	Upgrade Hub	1100	Yes
14	Binh Trung	Can open	Not open yet	No open		
15	Tan Huong	Can Upgrade	800	Upgrade Hub	1200	Yes
16	Binh Duc	No	700	Keep	700	Yes
Total			12/16 hub used		10/16 hub used	

References

Vietnam Department of Science and Technology. (2010) *Extending and developing Viet Nam sapote*.

Chopra, S. and Meindl, P. (2004) *Supply Chain Management: Strategy, Planning and Operation*. N.J: Prentice Hall.

Melkote, S. and Mark, S.D. (2001) An integrated model of facility location and transportation network design. *Trans. Res,* **Part A 35**, 515-538.

Taylor, B.W. and Russell, R.S. (2011) *Operations Management – Creating value along the supply chain.* John Wiley and Sons, NJ.

Loc, V.T.T. and Son, N.P. (2011) Value chain analysis of rice product in Mekong Delta. *Vietnam Science Journal,* **19**, 96-108.

Meng, Q. and Wang., X. (2010) Intermodal hub-and-spoke network design: Incorporating multiple stakeholders and multi-type containers. *Transportation Research,* **Part B 45**, 724–742.

Ishaf, R. and Sox, C.R. (2011) Hub location-allocation in intermodal logistics network. *Eur. J. Oper. Res*., **210**, 213-230.

Ishaf, R. and Sox, C.R. (2011) Design of intermodal logistics networks with hub delays. *Eur. J. Oper. Res*., **220**, 629-641.

Li, G., Hu, D. and Su, L. (2013) The model of location for single allocation multimodal hub under capacity constraint. *Procedia-Social and Behavioral Science,* **96**, 351-359.

George, O.W. and William, G.T. (1975) The Multiperiod Location-Allocation Problem with Relocation of Facilities. *Management Science* **22**, 57-65.

Sina, R., Mostafa, S., and Hossein, K. (2015) Capacity selection for hubs and hub links in hub location problems*, International Journal of Management Science and Engineering Management*.

Vi, H.N.B. (2020) Locating Hubs for Agricultural Product: A Case Study of Sapote in Chau Thanh District, Tien Giang Province, Vietnam. *International University-VNU, School of Industrial Engineering & Management.*

http://tiengiang.gov.vn/chi-tiet-tin?/sapo-mac-bac-kim-son. (Available access: May 20th 2021)

www.jsa.or.jp. (Available access: May 20th 2021).

MEDICAL WASTE COLLECTION VEHICLE ROUTING PROBLEM: A CASE STUDY IN THU DUC CITY

Tong Hong Phuc [1] and Ha Thi Xuan Chi[2]

[1] *Tong Hong Phuc, International University, tonghongphuc2411@gmail.com*

[2] *Ha Thi Xuan Chi, International University, htxchi@hcmiu.edu.vn*

Abstract. Medical waste collection vehicle routing problems have drawn great attention from academic researchers and practitioners because of the huge amount of operating costs from various healthcare services, especially in urban areas. This study attempts to minimum total traveling costs for medical waste collection in 11 wards of Thu Duc City, Ho Chi Minh City considering multiple disposal sites and drivers' rest periods. The mixed-integer programming model for vehicle routing problem with time windows (WCVRPTW) is applied to find exact solutions and The Nearest Greedy (NG) algorithm is employed to solve the large-scale problem. The solutions obtained by the NG algorithm are highly competitive with 300 times faster in terms of computation time and a slight gap of 8.81% with regards to average objective value compared to the exact ones.

Keywords: Medical waste collection, Vehicle routing problem, Nearest Greedy Algorithm, Drivers' rest periods.

1. Introduction

The vehicle routing problem with time windows (VRPTW) is considered as one of the best-known and the most important optimization problems in operational research, transportation science and computer science. It concerns the determination of the optimal set of routes for a group of vehicles with limited capacity. These vehicles always depart from a central depot and serve a set of customers with known demands and specific time windows and finally return to the central depot within a particular time window. Time window is the period between a start and end time of a customer that vehicle is allowed to start their service. Each customer is visited only one time by exactly one vehicle within the related time window.

There has been a large number of previous studies related to waste collection vehicle routing problem (WCVRP) using the heuristic techniques to solve the rising issue. Generally, there are two types of heuristic: (1) constructive heuristic algorithms, and (2) iterative improvement heuristic algorithms. A constructive heuristic algorithm refers to the technique that starts with an empty solution and repeatedly extends the current solution until a complete solution obtained. However, the iterative improvement heuristic algorithm, which is also known as "metaheuristic", operates by improving an initial solution to get the better one. This paper applied NG, one of the constructive heuristic algorithms to construct several initial solutions for WCVRP.

The rest of this paper is organized as follows: Section 2 reviews the relevant literature concerning the greedy algorithm in various applications. Section 3 presents the dataset and the technique employed to solve the problem raised. Meanwhile, the results of both sample size problems and the large size problem are discussed in Section 4. Lastly, some final remarks on conclusion and several suggestions for future work are presented in Section 5.

2. Literature review

Current research on WCVRP focuses more on improving the efficiency of collection practices by minimization of distance, logistic costs and travel time. Some researchers are trying to reduce the cost by improving the vehicles routes, other researchers are finding the most suitable location of disposal facilities and minimizing the number of vehicles used. Kim et. al (Byung-In Kim, 2006) studied the WCVRPTW and considered the four main objectives including number of vehicles, total travel time, route compactness and workload balancing. They solved the problem by applying two methods including an extension of Solomon's well-known insertion algorithm and a capacitated clustering-based algorithm. Ombuki-Berman et al. (Beatrice M. Ombuki-Berman, 2007) studied the similar problem by using a multi-objective genetic algorithm (GA) on the set of benchmark problems proposed by Kim et al (Byung-In Kim, 2006) . For the particular waste collection problem, this was the first study adopted GA approach. Although this method could solve the large-scale practical problem, the objectives related to route compactness and route balancing had not considered. Benjamin and Beasley (A.M. Benjamin, 2010) proposed a greedy constructive algorithm in order to generate initial solutions and developed three meta-heuristic algorithms: tabu search (TS), variable neighborhood search (VNS) and variable neighbourhood Tabu search (VNTS), which used the same benchmark as Kim et al (Byung-In Kim, 2006) . The results improved compared to the former of Kim et. al (Byung-In Kim, 2006) by using a disposal facility positioning procedure to evaluate routes. Buhrkal et al. (Katja Buhrkal, 2012) presented a mixed-integer linear programming (MILP) model for the WCVRPTW and proposed an adaptive large neighbourhood search algorithm for solving the problem. They tested their Adaptive Large Neighborhood Search (ALNS) algorithm on benchmark problems proposed by Kim et. al and on real-life instances provided by a waste collection company. The ALNS algorithm of Buhrkal et al. (Katja Buhrkal, 2012) performed better than the algorithms proposed by Benjamin and Beasley (A.M. Benjamin, 2010) . Recently, Han et. al (Hui Han, 2015) analyzed different methods and techniques to solve the Waste Collection Vehicle Routing Problem.

The number and size of healthcare facilities increased because of increasing the population. Various healthcare services are established to meet the demand and quality. Many previous researches have investigated healthcare waste management in healthcare facilities and institutions. In 2001, Li-Hsing Shih (Li-Hsing Shih, 2001) developed a computer system to solve an optimization problem based on a two-phased approach in terms of routing and scheduling systems for infectious waste collection. Then, Zkeik Hajar (Zkeik Hajar, 2018) focused on determining routes for healthcare waste on-site collection. The collection process is from points of generating the waste to its temporary storage area. with minimizing total traveling time and minimizing the transportation exposure risk. Masoumeh Taslimi (Masoumeh Taslimi, 2020) conducted an inventory routing schedule on a weekly basis to transport medical wastes to treatment sites under two objectives: minimize transportation risk and minimize occupational risk relevant with the temporary storage of hazardous wastes at the healthcare centers.

Most recently, Ziyuan Liu (Ziyuan Liu, 2020) proposed a VRP model for collecting medical waste between hospitals and temporary storage areas during COVID-19 by utilizing Q-value method to allocate medical waste transport vehicles and applying the immune-based ant colony algorithm, together with the tabu search algorithm, to arrange the correct pathways of waste transportation. Eventually, a complete “build–match–transport” system model for medical waste is established during these procedures. The application of this approach to the epidemic condition in Wuhan has generated excellent results, which has practical value and enlightenment to the emergency reaction and dispatch of Wuhan and other major cities.

3. Methodology

3.1 Mathematical model

PARAMETERS

$V=\{0\}\cup V^f \cup V^c$	Set of nodes
$A=\{(i,j) \mid \forall i,j \in V, i \neq j\}$	Set of arcs
G = (V, A)	A graph
m	Number of disposal sites
$V^f = \{1,\ldots,m\}$	Set of m disposal sites
n	Number of customers
$V^c = \{m+1,\ldots,m+n\}$	The set of n customers
$K = \{1,\ldots,nk\}$	Set of vehicles
t_{ij}	Travel time associated with arc (i,j)
s_i	Service time at node i
$[a_i, b_i]$	Time window at node i
s_u	Duration for break time (how long the vehicle/driver is idle)
$[a^u, b^u]$	The single break (rest) time window so that the rest period must start at the time within this interval
q_i	The amount of waste picked up at a customer $i \in V^c$
C	The capacity of the vehicles, so a vehicle filled to this capacity has to be emptied at a disposal facility before any other customer can be visited
R	The route capacity (total amount collected at customers for each route)
w_{ik}	The start time of service at node $i \in V$ for vehicle $k \in K$

(0,0')	The depot is split in a start and an end depot (0,0')
d_{ik}	The accumulative demand at node $i \in V$ for vehicle $k \in K$
M	A big parameter

VARIABLES

x_{ijk} $\begin{cases} = 1 \text{ if the vehicle } k \in K \text{ uses arc } (i,j) \in A \\ = 0 \; otherwise \end{cases}$

y_{ijk} $\begin{cases} = 1 \; if \; the \; break \; was \; taken \; between \; visitin \\ node \; i \; and \; j \; by \; vehicle \; k \\ = 0 \; otherwise \end{cases}$

MATHEMATICAL MODEL

Objective function: Minimize total travel cost.

$$\text{Min} \quad \sum_{(i,j)\in A} \sum_{k\in K} c_{ij} x_{ijk} \quad (1)$$

The objective function (1) is to minimize the total travel cost of all vehicles.

$$\sum_{j\in V} x_{0jk} = 1 \qquad \forall k \in K \qquad (2)$$

$$\sum_{i\in V} x_{i0'k} = 1 \qquad \forall k \in K \qquad (3)$$

Constraints (2) and (3) ensure all vehicles must leave and return to the depot respectively.

$$\sum_{i\in V} \sum_{k\in K} x_{ijk} = 1 \qquad \forall j \in V_c \qquad (4)$$

Constraints (4) imposes that each customer is visited exactly once.

$$\sum_{i\in V} x_{ijk} = \sum_{i\in V} x_{jik} \qquad \forall j \in V_c \cup V_f, k \in K \qquad (5)$$

Constraints (5) ensures that if the vehicle arrives at a stop, it must leave the stop.

$$a_i \le w_{ik} \le b_i \qquad \forall i\in V, \forall k\in K \qquad (6)$$

Constraint (6) specifies that the start time of service at node *i* must be in the time window of node *i*.

$$\sum_{i\in(0,0')} d_{ik} = 0 \qquad \forall k \in K \qquad (7)$$

Constraints (7) show that the vehicles must be empty when they start the route and when they return to the depot.

$$d_{ik} + q_i \le d_{jk} + (1 - x_{ijk})M \qquad \forall i \in V \backslash V_f, j \in V, k \in K \qquad (8)$$

Constraints (8) in place to make sure that the collected waste reflects the correct incremental volume of a particular stop when a vehicle visits that stop.

$$d_{ik} \le C \qquad \forall i \in V, k \in K \qquad (9)$$

Constraints (9) keep the amount of garbage collected at each stop within the vehicle capacity.

$$\sum_{(i,j)\in A} y_{ijk} = 1 \qquad \forall k \in K \qquad (10)$$

Constraints (10) impose that each vehicle driver must take exactly one break from the beginning to the end of one trip.

$$y_{ijk} \le x_{ijk} \qquad \forall (i,j) \in A, k \in K \qquad (11)$$

By constraints (11), a rest period of vehicle drivers happens once between node *i* and node *j* if they are connected.

$$w_{ik} + s_i + y_{ijk} s_u + t_{ij} \le w_{jk} + (1 - x_{ijk})M \qquad \forall (i,j) \in A, k \in K \qquad (12)$$

Constraints (12) make sure that the starting time of vehicle *k* at node *j* must be equal or larger than the starting time of vehicle *k* at node *i* plus service time at node *i* and travelling time from *i* to *j* as well as the break time of the driver.

$$a_u + s_u + t_{ij} \le w_{jk} + (1 - y_{ijk})M \qquad \forall (i,j) \in A, k \in K \qquad (13)$$

$$w_{ik} + s_i + t_{i,j} + \le b_u + (1 - y_{ijk})M \qquad \forall (i,j) \in A, k \in K \qquad (14)$$

Constraints (13) and (14) ensure that the break happens only within the time window. In other words, this constraint helps equalize the workload of each vehicle on each route.

$$x_{ijk} \in \{0,1\} \qquad \forall (i,j) \in A, k \in K \qquad (15)$$

$$y_{ijk} \in \{0,1\} \qquad \forall (i,j) \in A, k \in K \qquad (16)$$

$$d_{ik} \geq 0 \qquad \forall i \in A, k \in K \qquad (17)$$

Constraints (15), (16) and (17) impose binary conditions, nonnegative integer constraint, and nonnegative conditions on the variable set.

3.2 Using Nearest Greedy algorithm

In this study, the NG approach had been used to construct vehicle routes in order to solve an actual medical waste collection problem. At the begining, the route started from node 0. Then, the NG algorithm was employed to determine the closest customer from node 0 to be served by comparing all distances from nodes of 0 to each of n customers. After that, the algorithm selected the node with the shortest distance, adding to the present route, and updated the remaining number of customers to be served. After each customer had been served, the algorithm checked vehicle capacity. If the vehicle had been fully loaded, it would need to dispose the waste at one of the two disposal facilities. At that step, the algorithm continued to make comparison between the distance from current node to go to disposal site 1 and to go to disposal site 2. The process continued until all customer nodes had been served. Before returning to node 0, the vehicle would have to go to the disposal site to empty its load. The algorithm calculated the total travel cost travelled by the vehicle and displayed the constructed vehicle route. The detail flow chart of the NG algorithm applied in this paper is shown in Figure 1.

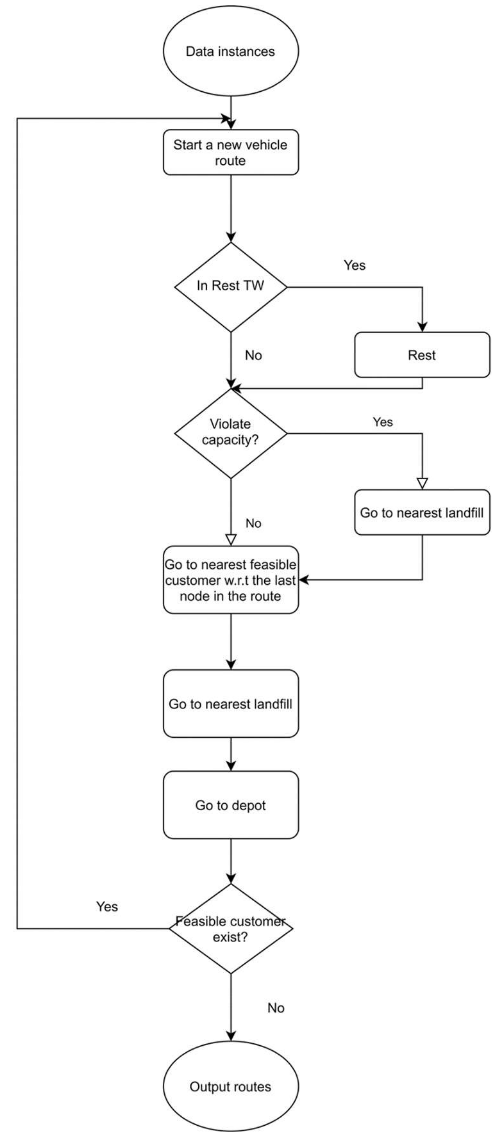


Figure 1: The detail flow chart of Nearest Greedy algorithm

4. Case study:

In this study, medical waste data are collected from former Thu Duc district Ho Chi Minh City, Vietnam with one depot (denoted as D), two disposal sites (denoted as DS), and 150 customers (denoted as C). Sample of demand (kilogram), locations, needed service times (second) and service time window (second) of customers (see Table 1). The collection time are from 8:00 to 16:00. Each driver has a one-hour break between 11:00 and 13:00. There are 5 motorbikes totally. The capacity of each motorbike is 30 kilograms since customer demands are relatively small. When collecting waste, each vehicle starts and ends at the depot with zero waste capacity. Each customer is served exactly once by one vehicle. The total demand served on the trip of a vehicle cannot exceed the vehicle's capacity. All customers must be served within their time windows. Vehicles must visit them between the earliest service time and the latest service time. If the vehicle arrives earlier than customer the earliest service time, it must wait until the earliest service time to start collecting waste.

Table 1: Depots, disposal sites and customers sample dataset

No.	Address	Service time	Demand	Earliest service time	Latest service time
D	11 Khổng Tử, Bình Thọ	-	-	0	28800
DS 1	234 Tam Bình, Tam Phú	300	-	0	21600
DS 2	64 Lê Văn Chí, Linh Trung	300	-	0	21600
C1	224-226 Võ Văn Ngân, Bình Thọ	150	1.5	0	21600
C2	12 Lam Sơn, Linh Tây	150	2.5	0	21600
C3	919 Kha Vạn Cân, Linh Tây	150	3.5	0	21600
C4	121 Tô Ngọc Vân, Linh Tây	150	8.3	0	21600
C5	Số 19 Công Lý, Bình Thọ	150	7.5	0	21600
C6	42 Đặng Văn Bi, Bình Thọ	150	6.8	0	21600

5. Results

5.1 For small-size problems

In this paper, the exact mathematical model was generated by CPLEX while the NG algorithm was run by Python language. As can be seen in Table 2, for small size problems (from 4-14 customers), CPLEX obtains better objective values than the NGalgorithm. The gap of objective value between two methods is larger when increasing sample size. The average gap of optimal objective value is 8.81%. Moreover, the results generated by Python have a significant advantage in computation time which only comes up with the solution in 1 second. In small-scale instances, the NG heuristic can obtain the optimal solution in a very short time while CPLEX takes nearly one hour to obtain solutions.

Table 2: Summary results solved by CPLEX and Python in terms of objective and computation time

No of nodes	Optimal values		Computation time		Optimal value gap	Computation time gap
t	ZC	ZP	TC	TP	(ZP-ZC)/ZP	TC/TP
5	9,4	9,730	3	1	3.4%	3
6	10,40	11,07	3	1	6.1%	3
7	13,91	15,00	4	1	7.3%	4
8	15,45	16,99	4	1	9.1%	4
9	26,50	28,63	4	1	7.4%	4
10	26,28	28,96	14	1	9.3%	14
11	26,38	29,03	38	1	9.1%	38
12	25,9	28,95	181	1	10.5%	181
13	25,91	29,58	1200	1	12.4%	1200
14	26,21	30,31	1580	1	13.5%	1580

5.2 For a large-size problem

With 150 customers and 5 vehicles, by applying the NG heuristic, the routes of vehicles and the place for drivers take a rest are shown in Table 3.

Table 3: Customer sequence of NG heuristic for large size problem

Vehicle number	Route	Rest node
Vehicle 1	[0, 144, 20, 5, 130, 68, 42, 6, 145, 48, 150, 1, 92, 14, 8, 22, 121, 124, 123, 40, D2, 21, 16, 93, 147, 98, 50, 126, 122, 128, 59, 66, 71, 90, 140, 65, 78, D1, 0]	[121]
Vehicle 2	[0, 58, 61, 60, 63, 10, 11, 2, 52, 13, 41, 37, 29, 3, 47, 4, 19, 12, 106, D2, 118, 97, 43, 114, 96, 18, 105, 117, 125, 53, 127, 54, 143, 149, 62, 102, 75, 69, 151, 132, 137, 138, D1, 0]	[114]
Vehicle 3	[0, 32, 35, 51, 115, 111, 38, 112, 133, 25, 142, 139, 113, 131, 141, 100, 64, 17, 151, 134, 28, 85, 27, 31, 46, 45, 76, 103, 135, 129, 77, 148, 67, 79, 73, 72, D1, 0]	[64]
Vehicle 4	[0, 7, 44, 30, 116, 104, 99, 39, 34, 119, 80, 107, 15, 94, 74, 101, 24, 87, 36, 56, 84, 152, 57, 136, 83, 86, 82, 70, 146, D2, 0]	[94]
Vehicle 5	[0, 26, 110, 23, 33, 109, 88, 49, 55, 81, 108, 120, 91, 95, 9, 89, D1, 0]	[9]

6. Conclusion

This study solves medical waste collection routing problem cconsidered time windows and the rest period of the driver by applying the NG Algorithm. For the small-size problem, the gap between the objective values of CPLEX and the Nearest greedy algorithm is 8.81%. However, the NG method comes up with the solutions in much faster computation time, 300 times faster than CPLEX. Moreover, it only takes around 8 seconds to generate the routes for a large-scale problem from NG algorithm. The effective solutions to this problem also produce significant economic benefits for the company. By implementing this solution, it will cost the company just 627,144 VND for the total traveled cost – the smallest amount of money. The method can utilize all available resources of the company. Finally, medical waste collection is a laborious, difficult and risky task since the waste may be infectious or hazardous. In this study, the driver rest period is appropriately taken into consideration, which means the drivers can take a break during their workday and help retain their productivity, mental well-being and overall work performance. It is recommended that the time windows of customers can be divided into groups which have the same interval of time. One vehicle can serve one group of customers.

References

Byung-In Kim, S. K. (2006). Waste collection vehicle routing problem with time windows. *Computers & Operations Research*, 3624-3642.

Beatrice M. Ombuki-Berman, A. R. (2007). Waste collection vehicle routing problem with time windows using multi-objective genetic algorithms. *Computational Intelligence.*

A.M. Benjamin, J. B. (2010). Metaheuristics for the waste collection vehicle routing problem with time windows, driver rest period and multiple disposal facilities. *Computers & Operations Research*, 2270-2280.

Hui Han, E. P.-C. (2015). Waste Collection Vehicle Routing Problem: Literature Review. *Promet – Traf c&Transportation*, 345-358.

Katja Buhrkal, A. L. (2012). The Waste Collection Vehicle Routing Problem with Time Windows in a City Logistics Context. *Procedia - Social and Behavioral Sciences,* 1877-0428,

Li-Hsing Shih, H.-C. C. (2001). Routing and Scheduling System for Infectious Waste Collection. *Environmental Modeling & Assessment*, 261–269.

Masoumeh Taslimi, R. B. (2020). Medical waste collection considering transportation and storage risk. *Computers & Operations Research.*

Ziyuan Liu, Z. L. (2020). Path Optimization of Medical Waste Transport Routes in the Emergent Public Health Event of COVID-19: A Hybrid Optimization Algorithm Based on the Immune–Ant Colony Algorithm. *Environmental Research and Public Health.*

Zkeik Hajar, D. B. (2018). Onsite medical waste multi-objective vehicle routing problem with time windows. *International Conference on Logistics Operations Management* (GOL), 1-5.

AN APPLICATION OF SIMULATION AND OPTIMIZATION TO OUTPATIENT APPOINTMENT SCHEDULING SYSTEMS - A CASE STUDY OF HEALTH CARE CLINIC IN HO CHI MINH CITY

Le Dinh Hoang Dang[1], [*]Nguyen Hoang Huy[2]

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: [1]ldhd0169@gmail.com, [2]nhhuy@hcmiu.edu.vn

Abstract. The paper introduces the use of discrete event simulation modelling to improve working process of an orthopedic clinic in Ho Chi Minh City. In service-related businesses, long waiting time is the one of the most critical factors of achieving customers' satisfaction. Healthcare service, in particular, is also facing delay in service due to the increases of demand. Therefore, besides public hospitals, private clinic is a solution for patients to easily access healthcare with the uses of early booking. In another perspective, with the demand increases over time, long waiting time in inevitable. The studied private clinic is shown to be effective in treating patient, but the dissatisfaction comes from long waiting time creating bad service to the customers and staffs' morale issues. The study aims at identifying the causes of the problem and propose an optimize solution. Simulation models has proved to produce an approximate current state model of the clinic and detail about where the issues occur. Based on the result of the simulation model, an optimize tool was applied to produce an efficient appointment scheduling which show significant reduction in waiting time of patients.

Keywords: Health care clinic, Discrete event simulation, Appointment Scheduling.

1. Introduction

1.1. Background Overview

With the continuous increases in healthcare demand, public hospitals are currently facing congestion in patient flow throughout service time. As a result, patients with busy schedule are commonly visiting hospital outpatient clinics. These clinics typically offers early appointment booking service, and patients are expecting to have their service begin on time. However, patients still end up waiting even though they have reserved time slots. In reality, it is common for clinics to book multiple patients simultaneously to ensure they are always available to see doctors or other expensive, scarce health care resources. However, because patients have a tendency to arrive earlier than their appointments, long waiting time is inevitable.

Waiting has always been a problem causing dissatisfaction in the service quality of outpatient clinics. In another perspective, non-attendance at outpatient clinics is a significant problem in healthcare service. It is related to the prolonged waiting times from referral to appointment. Non-attendance is an undesirable occurrence where it can be avoided. This suggests that for health care systems where patients choose a health care provider, shorter waiting times are a competitive advantage.

This study will focus on observing the patient flow at Ho Chi Minh City with simulation method. This method has proven to be successful in analysing and improving operating systems in industry settings, including the healthcare industry. Simulation tools have improved, and an emphasis on cost control and efficiency has become more critical in healthcare. Simulation also provides various alternative patient pathways with the help of animation and visualization.

1.2. Problem statement

By studying many articles, here are the conclusion of what outpatient clinics are currently facing:

- Long waiting time: With the rise in demand, most patients have to wait up to an hour or more in order to meet their appointment despite arriving on time. This affects not only the patient flow of the clinics but most importantly the satisfaction of the patients when using healthcare service.
- High utilization of doctors: The negative effect of waiting is not only on the patients waiting for service but also on the clinic's staff due to the growing lines of patients, put staffs under hard pressure of having to deal with work and patient complaints at the same time. In the long run, such pressure can create morale problems and likely contributes to absenteeism.

1.3. Objectives

The main objective of this paper is to show how a structured patient flow analysis can significantly improve the service quality of the clinic by minimizing the average waiting time and consider two sub-objectives, which are improving staff morale and patient scheduling procedure to produce significant improvement to the clinic's performance and customers' satisfaction.

1.4. Scope and limitations

This paper aims to propose a simulation method combine with appointment scheduling in order to minimize the waiting time of patients in outpatient clinics. Due to the limitation of time, the study will cover one clinic only, and collected data is only through direct observation and information provided by staffs at the clinic.

2. Literature review

2.1. Patient flow in healthcare service

Many hospitals around the world have proposed solutions to overcome the long waiting time problem in outpatient clinics such as hospitals in Chicago, China. The following paragraphs will explain their solutions and the progress of their improvements. In the sample Chinese Hospital (Shafer 2010), three models were presented to demonstrate how different strategies can be used to improve the performance of the outpatient patient flow. In Model A, ''doctors sharing patient'' any available doctor could see any patients from the system in the first examination in which the waiting time has decreased. However, results shown that increases in waiting time were found in other areas. In Model B, ''adding volunteers'', five volunteers were introduced to the system, which lowered the waiting time significantly. In Model C, ''changing volunteer priorities'', changing the location of volunteers dynamically, an improvement in resource utilization was achieved. The best reduction they have got was 78%. A simulation study of an Ears, Nose and Throat clinic at the University of Illinois Medical Center, Chicago, USA was conducted by (Haji and Darabi 2011) in order to reduce the total patient waiting time in the system. Simulation showed that changing the appointment schedule policy is more effective than adding an extra resource to the system. In the first scenario, one resident doctor has been added to the system to serve all patients' types. This solution achieves a 5.29% reduction in waiting time. The second scenario has changed the appointment policy which succeeded to attain a 15% improvement.

These solutions depend mainly on adding more human resources or changing business or management policies to improve waiting time. This may not be sufficient as the applied software services in outpatient clinics need to be addressed as well. The main aim of this article is to study the current patient flow from a software perspective to reduce the waiting time while utilizing the currently available resources.

2.1. Discrete-event simulation (DES) in healthcare

DES has been shown to be effective in many healthcare settings. Some models of outpatient clinics aim to improve patient flow, reduce waiting time, maximize staff utilization. These outpatient models are tested through changes to patient scheduling, patient routing, and internal work processes. The majority of discrete-event simulation studies that focus on patient scheduling and admissions are focused on outpatient clinics. (Wagner and West 2004) outline a simulation framework for analyzing scheduling rules for outpatient clinics. This framework addresses four key components of an outpatient clinic scheduling system: demand for appointments, supply of physician time blocks, patient flow, and the scheduling algorithm. The study provides a demonstration of the framework for a pediatric ophthalmology clinic and discusses some challenges for adapting the framework to other settings. Ferrin, Miller, and McBroom 2007) demonstrates a DES model that is used to develop processes for increasing throughput in an emergency department (ED) in the USA as part of a system that permitted the diversion of ambulances in peak demand periods, which had financial implications for the client ED. Their investigations included the introduction of discharge lounges, shortening the length of stay, and bypassing triage. (Parks et al. 2011) shows how discrete-event simulation was used in an adult medicine clinic within a large, tertiary care, academic medical center. Results of the simulation indicated that system bottlenecks were present in the medication administration and check-out steps of the clinic process. The simulation predicted that matching resources to excessive demand at appropriate times for these bottleneck steps would reduce patients' mean time in the system. In the case of an orthopedic clinic, (Rohleder et al. 2011) applied DES with the effort to identify improvement alternatives including optimized staffing levels, better patient scheduling, and an emphasis on staff arriving promptly. As a result, statistical analysis of data taken before and after the implementation indicate that waiting time measures were significantly improved and overall patient time in the clinic was reduced.

2.3. Appointment scheduling

Appointment scheduling in an outpatient clinic consists of determining the sequence of patients to optimize some performance measures. Many papers can be found in the literature on appointments scheduling. They can be divided into two categories: static and dynamic problems. When appointments are scheduled without the possibility to modify the schedule, it is considered as a static problem. This is how appointments are scheduled in most healthcare services. Therefore, many papers are devoted to these scheduling appointment problems.

When the appointment schedule can be constantly reconsidered during a given day depending on the current state of the clinic, it is considered as a dynamic problem.

Most appointment schedules are constructed based on specific decision rules to minimize negative effects related to patients arriving late, no-shows or walk-ins. In 2003, (Cayirli and Veral 2003) reviews the literature dealing with outpatient scheduling in health care. Studies had analyzed appointment systems for effectively regulating the flow of patients so that both patient waiting times and doctor idle times are minimized. The most studied rules are:

- Variable blocks/fixed intervals rule consisting of planning appointment periods of varied size (one, two or more patients) while keeping fixed intervals.
- Individual blocks/variable intervals rule consisting of planning appointment blocks of size one with intervals of varied size.
- DOME rule consisting of planning m appointment blocks during fixed short period of time (intervals) at the beginning and at the end of the day, and the remaining ones during fixed long periods of time in the middle of the day.

In relation to simulation, (Klassen and Yoogalingam 2009) proposed a simulation optimization approach to determine optimal rules for a stochastic appointment scheduling problem in an outpatient clinic. Results show that the dome scheduling rule proposed in prior literature is robust over many different performance measures and scenarios. In another case, (Berg et al. 2010) showed that there is a positive linear relationship between the number of patients examined and the number of examination rooms. They tested ratios of one examination room per doctor (1:1) and those of two examination rooms per doctor (2:1). With a higher ratio, more auxiliary personnel (nurse, secretaries) are needed to improve performance. In recent years, (Baril, Gascon, and Cartier 2014) uses discrete event simulation to study the relationships and interactions between patient flows, resource capacity (number of consulting rooms and number of nurses) and appointment scheduling rules in order to improve an outpatient orthopedic clinic performance. An experimental design was developed to test how to assign consulting rooms and nurses to each orthopedist considering four appointment scheduling rules and three patient flow types of varying complexity. As a result, the author concluded that the above factors must be adapted to different patient flows.

Finally, all these appointment scheduling problems are either solved with analytic or with simulation techniques. Considering that this study includes many factors interacting with each other, a possible approach is using discrete event simulation to test scheduling rules. Very little cases where the dome rule is adopted to solve scheduling problem, however, due to the fluctuation in patients' arrival time, the dome rule has some potential scheduling in healthcare environment.

3. Methodology

3.1. Clinic operation

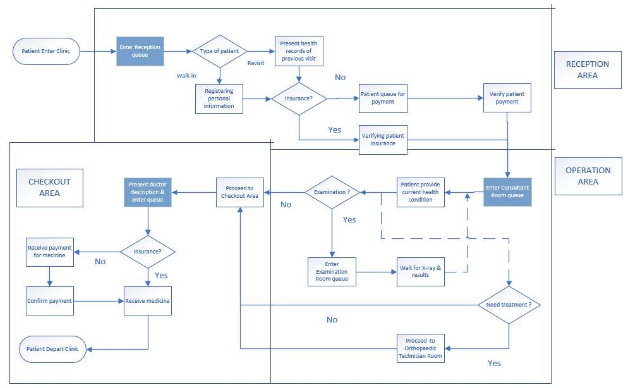


Figure 1. Clinic operation system

The considered orthopedic clinic is a private clinic serving around 70-100 patients per day comparing with the sample size of 50-100 gathered in (Rohleder et al. 2011), which is considered enough to conduct a simulation research. The cast clinic is an outpatient clinic which accommodates the majority of outpatient orthopedic patient appointments for the hospital including fractures, some spinal injuries, hand and wrist disorders.

The clinic generally open from 8 am to 4 pm Monday through Saturday and is closed on holidays. On most days five doctors are scheduled at the clinic, however, some may be scheduled depending on holidays, vacations. The clinic operations include:

- 5 Consultant rooms
- 2 Receptionist to check-in patients when they arrive and schedule appointments.
- 2 x-ray machines
- 2 x-ray technicians
- 2 orthopedic staffs
- 2 payment staffs
- 2 medical staffs
- 1 register staffs

In the models, there will be some patients in progress at the end of the day, and the clinic will arrange them with another appointment in the next day.

3.2. Data collection

Four primary performance measures were chosen to quantify the impact that any changes to the clinic's processes might have:

1) The total time a patient spends in the clinic

2) The time a patient waits for an x-ray
3) The time a patient waits for the consultant
4) The percentage of patients whose clinic visits were completed in 60 minutes or less

Through observing at the clinic patients although arrive early than their appointment still has to wait from 20-30 minutes until they meet the consultant. As for the x-ray area, a significant waiting time of 10 to even 20 minutes of waiting for each patient.

These four measures clearly relate to patient satisfaction with waiting and clinic congestion. In particular, measure 4) was included because Clinic administrators were given a target of getting all patients through the clinic in 60 minutes or less. However, it was recognized that measures 1) and 4) were not completely under the control of the Clinic, as some patients were arriving very early for their appointments. As patient satisfaction is most directly related to how long a patient waits to be seen, a sixth measure, length of the initial wait in the waiting room, was added. As the patient arrival pattern is unlikely to change in the short run, a combined reduction in this measure and in measures 1) and 4) would indicate that system performance has truly improved and it is not simply that patients are arriving less early to their appointments. In the long run, process improvements would hopefully lead to shorter and more consistent waiting times that in turn would lead to less variability in patient arrival times.

3.3. Input data analysis

Table 1 shows the distribution of each department's processing time. In real time situations, we can often estimate the maximum and the minimum values, the values are most likely, even if we do not know the average value and the standard deviation. For this reason, triangular distribution will be chosen as the main distribution for the time data that this paper is targeting. Table 2, 3 and 4 provide detailed information about clinic staffs, working shift and number of arrivals in each period.

Table 1: Processing time distribution

Process name	Distribution	Parameter (minutes)
Register Counter	Triangular	TRIA (2,3,5)
Patient Reception Counter	Triangular	TRIA (2,4,5)
Service and Medication Payment Counter	Triangular	TRIA (2,4,5)
Consultant room (1-5) for Revisit, Walk-in, and follow up patient	Triangular	TRIA (10,15,20), TRIA (15,18,20), TRIA (10,12,15)
X-ray room (1-2)	Triangular	TRIA (5,10,15)
Technician room	Triangular	TRIA (20,30,40)
Medication counter	Triangular	TRIA (2,4,5)

Table 2: Resources of each department

Resource	Quantity	Department
Receptionist	2	Patient Reception Counter
Doctor	5	Consultant Room
Payment Staff	2	Payment Counter
Register counter	1	Register counter
Xray Staff	2	Xray room
Orthopedic Staff	2	Technician room
Medical Staff	2	Medication counter

Table 3: Available working time

Working time	Value
Number of shifts per day	2
Working time/shift (hours)	4
Break for lunch (hours)	1
Available working time (hours)	7

Table 4: Number of arrivals in each period

Time	Number of arrivals	Time	Number of arrivals
8h-8h29	60	12h-12h29	0
8h30-8h59	50	12h30-12h59	0
9h-9h29	56	13h-13h29	62
9h30-9h59	53	13h30-13h59	38
10h-10h29	43	14h-14h29	41
10h30-10h59	21	14h30-14h59	27
11h-11h29	14	15h-15h29	25

3.4. Model explanation

Firstly, each module in the simulation will be briefly described to give an overview on how the system works.

Patient arrivals time was collected and convert into arrival rate with time between arrival of 30 minutes.

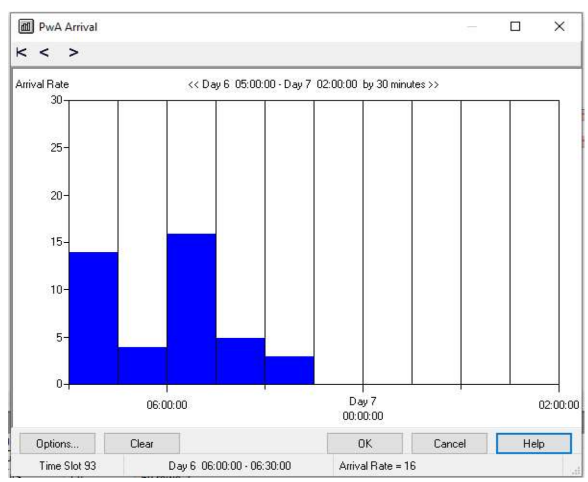


Figure 2. Arena Schedule for patients with appointment

A schedule for walk-in patient was built as well but with another create module to clarify the process to match reality, and they have to undergo a register procedure after entering Patient reception counter.

Both type of patient is then queue up according to their appointment and arrival time at the clinin. For each patient, there is a medical record which they have to present at Patient reception counter for Reception Staff to schedule them into which room. Patients with insurance can proceed to Operation area for their checkup, while noninsurance patients is guided to Payment booth 1 for service payment.

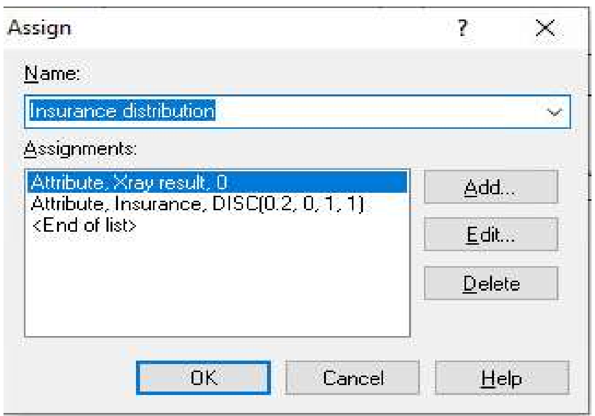


Figure 3. Assign module for Insurance check

The percentage of this module was collected base on observation with the outcome of 20 % of patient without insurance and 80% with insurance. By counting the number of patients which have to pay for their service at the clinic, the percentage can easily be achieved.

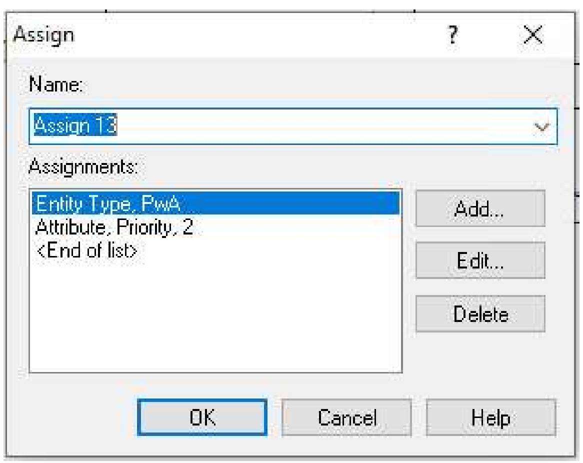


Figure 4. Assign module for Priority in queue

Assign module allow the model to identify the type of patient for processing time distribution as well as priority in queue. Patient with appointment is divided into two types which is on time and late arrival. Patient who is late for their appointment will have priority the same as walk-in patient.

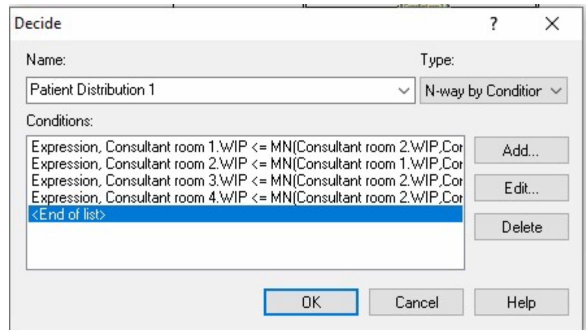


Figure 5. Decide module for distributing patient in Consultant room

This module will assign entities into available slots according to the minimum number of "Work in progress" patients between five rooms. Example expression of Patient Distribution module is described as follow: Consultant room 1.WIP <= MN (Consultant room 2.WIP, Consultant room 3.WIP, Consultant room 4.WIP, Consultant room 5.WIP). This expression will assign entities only if the number of patients in progress is the lowest of all five rooms otherwise it will choose another room.

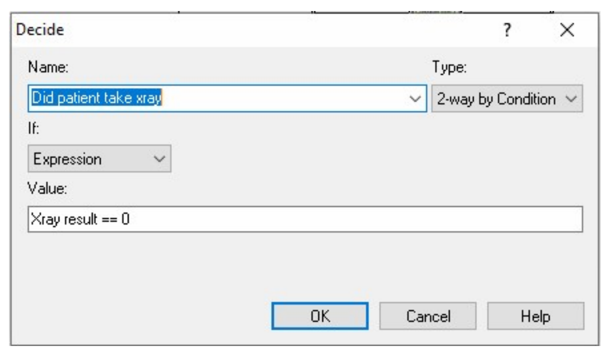


Figure 6. Decide module for distributing patient in Xray room

The module will decide base on the Attribute Xray result of each entity, which is assigned from beginning. Those with value 0 (have not taken Xray) will be moved to Xray area and assigned with value 1 (already taken Xray), also highest priority and shorter processing time. Those with Xray result value = 1 or higher will go back to join Consultant room queue but with higher priority than other entities through Route 7 module. Finally, doctors will decide whether or not patient need cast care in Technician room, and proceed to the Check out Area for medicine.

3.5. Data validation

3.5.1. Model validation

Model validation is an obligatory task that must be done to examine the differences between the model and the collected data. The result that is produced has to fall within an acceptable level of accuracy. The validation of the model is performed by comparing statistics on the processes with data collected by the sales department. Hypothesis testing was used in this step. In statistics, t-tests are a type of hypothesis test that allows you to compare means. They are called t-tests because each t-test boils your sample data down to one number, the t-value. The paired t-test and the one-sample t-test are the same tests in disguise. A one-sample t-test compares one sample mean to a null hypothesis value. A paired t-test simply calculates the difference between paired observations and then performs a one-sample t-test on the differences. The table below is the results from 10 replications with no warm-up period. Compare the average Number of Patients from Arena Simulation with the real Number of Patient data from reality, we get a table below:

Table 5: Data comparison

Day	Real Data	Model Data
1	76	79
2	74	76
3	82	85
4	79	82
5	85	90
6	91	80

In this paper, the one-sample t-test for paired data is used to validate the model:

The one-sample t-test for paired data is defined as:

$$H_0: \mu_D = \mu_1 - \mu_0 = 0$$

$$H_1: \mu_D = \mu_1 - \mu_0 \neq 0$$

Assume that:

μ_1: Average Number of Patients in Arena Simulation

μ_0: Number of Patient of real data

Mean of differences:

$$\overline{\mu D} = \frac{1}{n} \times \sum_{i=1}^{n} (\mu 1 - \mu 2)$$

Standard deviation of differences:

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (\mu D - \overline{\mu D})^2}{n-1}}$$

Standard error of differences:

$$SE = \frac{SD}{\sqrt{n}}$$

The value t statistic:

$$t = \frac{\overline{\mu D}}{SE}$$

Degree of freedom:

df = n – 1

Level of significant:

$\alpha = 0.05$

Lower bound and upper bound of confidence interval:

$$(\overline{\mu D} - SE \times t\text{df},\ \alpha/2;\ \overline{\mu D} + SE \times t\text{df}, \alpha/2)$$

Using the Microsoft Excel, the calculation of all the values is below:

Table 6: Difference between Arena simulation and Real data

Day	Real Data	Model Data	Difference
1	76	79	3
2	74	76	2
3	82	85	3
4	79	82	3
5	85	90	5
6	91	80	-11

Table 7: Data evaluation

Mean of differences	0.83
Standard Deviation of Differences	5.88
Standard Error of Differences	2.40
T Statistic of Differences	0.35
Degree of Freedom	5.00
95% Confidence Interval	2.57

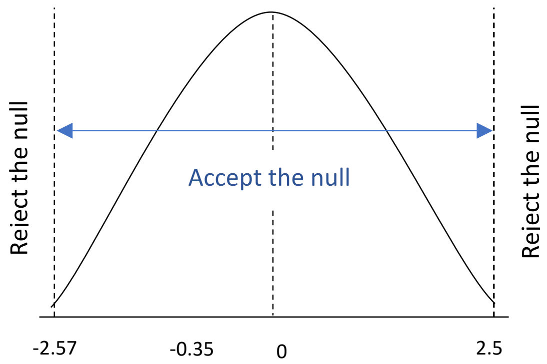


Figure 7. One-sample t-test for paired data

From the figure above, the t statistic is in the "accept the null" area, so we do not reject H_0 . Therefore, the hypothesis test shows that there is no significant difference between the actual data and the simulation model result.

3.5.2. Validating Input-Output Transformation

The goal of the validation process is twofold:

- To produce a model that represents true system behavior closely enough for the model to be used as a substitute for the actual system.
- To increase to an acceptable level, the credibility of the model, so that the model will be used by managers and other decision makers.

This value will be used as the mean value: μ = 487. The hypothesis testing:

- Ho: $\mu_0 \neq 487$ patients
- H1: $\mu_0 = 487$ patients
-

Table 8: Total number of patients in simulation model (10 replications)

Replication	Total number of patients
1	488
2	485
3	482
4	471
5	490
6	485
7	477
8	489
9	493
10	481
Level of significance	α = 0.05
Sample size	10
Standard deviation	9.89
Sample mean	484.00
T statistic	-1.39
Critical value	2.26
Upper bound	488.87
Lower bound	479.13

Since $|t_0| < t_{0.025,9}$, do not reject Ho.

Best case:

|Upper bound – 484| = |488.87 – 487| = 1.87 > 1

Worst case:

|Lower bound – 487| = |479.13 – 487| = 7.87 > 1

We can say that addition replications are needed to reach the decision.

3.5.3. Determine number of replications

According to *Simulation with Arena 2th*, we can easily collect statistical data after running the simulation on the Arena by making some number n of independent replications.

The most common number of replications mentioned is 10 replications. This provides enough replications to achieve reasonable statistical confidence, which then adds a larger number of replications.
After running the Arena model with 10 replications with the key performance indicator is number of patients served.

Table 9: Number of Patients served per replication.

Replications	Number of patients served
1	488
2	485
3	482
4	470
5	490
6	485
7	477
8	489
9	493
10	481

It also can determine the 95% confidence interval half width as the figure below:

Number Out	Average	Half Width	Minimum Average	Maximum Average
PwA	388.80	19.34	354.00	437.00
WP	82.2000	7.44	65.0000	96.0000

Figure 8. The output of entity patients served.

This statistic helps determine the reliability of the results from the replications. The half width value is defined that 95% of the average of replications of samples to be determined within the range where occur error is the average ± half width. Therefore, with the 10 original replications, the average 95% confidence interval for the number of patients served was: 471 ± 26.78.

Table 10: Statistical analysis of 10 replications.

	Value
Sample Mean	471
Sample Standard Deviation	9.89
95% Confidence Interval Half Width	26.78
Minimum Summary Output	419
Maximum Summary Output	533

In order to calculate the replications that the model needs to run, an appropriate level of error must be given. Based on *Simulation with Arena 2th* this level of error will be less than or equal to one-third of the calculated 95% confidence interval of half width. There are two formulas to calculate the replications:
In this situation, the 95% Confidence Interval Half Width is 26.78, the value we should choose is 8.9, but let choose a smaller value h = 3 to achieve better results

First approximate formula based on standard deviation:

$$n \cong z^2_{1-\frac{\alpha}{2}} \frac{s^2}{h^2}$$

$$n \cong z^2_{1-\frac{\alpha}{2}} \frac{s^2}{h^2} = 1.96^2 * \left(\frac{9.89^2}{3^2}\right) = 42.68$$

Second approximate formula based on half width:

$$n \cong n_0 \frac{{h_0}^2}{h^2}$$

$$n \cong n_0 \frac{{h_0}^2}{h^2} = 10 * \left(\frac{26.78^2}{3^2}\right) = 89$$

It can be seen that the second formula is bigger so 89 replications will be the minimum acceptable number of replications apply for this model. The final step is to rerun the model with new replication which is 100 replications to see the difference.

Number Out	Average	Half Width	Minimum Average	Maximum Average
PwA	387.45	3.62	354.00	437.00
WP	84.0100	1.92	65.0000	108.00

Figure 9: The output of entity patients served with 100 replications.

From the figure above, the half width has decreased significantly through 100 replications, the final number of patients served is 474.46 ± 5,54.

4. Result analysis

4.1 Result from simulation

After running 10 replications of the system, the model produced the following results:

Table 11: Waiting time per patient

Department	(minutes)
Check-in Area	
Patient Reception Counter	0.84
Register Counter	0.49
Service Payment Counter	0.42
Operation Area	
Consultant room 1	18.56
Consultant room 2	17.36
Consultant room 3	16.12
Consultant room 4	15.13
Consultant room 5	14.08
Xray machine 1	8.29
Xray machine 2	6.05
Technician room	21.69
Check-out Area	
Medication Counter	0.98
Medication payment	0.29

While the patients' paths vary significantly through all the clinic resources, by a significant margin, the x-ray machines/technicians and the surgeons are the busiest resources. The simulation model provide average waiting time of each patients in the clinic, as seen from Table 5.1 the Operation Area have the longest queue time out of the three area in the clinic. These congestions occur due to a large number of patients arrive at a certain time (commonly 8:00-9:30 a.m,). Two sections that need most attention are Consultant room and Xray machine, although Technician room also have long queue, it is acceptable due to the nature of work.

Table 12: Average Utilization by Operation Area

Resource staffs	Value
Xray staff 1	0.9296
Xray staff 2	0.8174
Doctor 1	0.9752
Doctor 2	0.9577
Doctor 3	0.9066
Doctor 4	0.8189
Doctor 5	0.7721
Orthopedic Staff 1	0.8436

4.2. Result from OptQuest

The objective of this optimization phase is to minimize the average total waiting time of patients in the clinic base on adjusting the rate of arrival (N: number of entities per arrival and T: time between arrival). Thus, the OptQuest model will focus on variable such as the average total waiting time of each entity, waiting time of entity spend in each room.

By adjusting N and T, the clinic will easily schedule their patient to maximize the performance of their staffs as well as minimize the waiting time of their patients. In the literature review, (Rohleder et al. 2011) [6] used N= 1, 2, 3 and T = 10, 15, 20, 30 to optimize the appointment of the clinic, but if N > 1 the time between two or more patient will be hard to control. Thus, in this approach, I will use N = 1 and vary T throughout the optimization process to give an exact range between arrival so the clinic can easily arrange their patients.

Parameters:

- PwA: patient with appointment
- WP: patient without appointment
- N: number of entities per arrival (N=1)
- T: time between arrival

Procedure:

- **Tool -> OptQuest for Arena -> New optimization**

Choose Controls (which is N and T in this case)

- N is set to 1
- T is set to discrete with lower bound of 1, suggested value of 2, upper bound of 10, with step of 1 (T will increase or decrease 1 min after each simulation)

Choose Responses

These are the values that are used in constraints and objective

Choose Constraints.

Consultant room 1-5.Waittime < 10 (minutes)

Xray machine 1-2.Waittime < 5 (minutes)

Choose Objective. (Minimize)

Pwa.Waittime + WP.Waittime

Optimizing process

After adding constraints and objective, OptQuest will then choose an initial solution set of N and T, which will be used to improve the solution. In addition, OptQuest will also show the infeasible solution and show which constraint did the solution violate. The detail on the improvement of all solution will be described by OptQuest in the below graph, showing multiple infeasible solutions until it is optimized.

By using OptQuest, optimize solution can be apply directly in the simulation without any rerun of the simulation model itself compare to those of CPLEX or heuristic approaches. The computational time of this is not long, around 2-5 minutes depending on how much replications are needed for each simulation.

The optimal rate of arrival is determined after 100 replications of ten T values, the result show that N = 1 and T = 6 is the most suitable solution in order to minimize the waiting time of patient. Below is the comparison of waiting time value of current and final state:

Table 13: Waiting time result from OptQuest

Responses	**Final state (minutes)**	**Current state (minutes)**
PwA.Waittime	11.32	47.85
WP.Waittime	11.44	50.2
Consultant room 1.Waittime	6.31	18.56
Consultant room 2.Waittime	4.89	17.36
Consultant room 3.Waittime	3.65	16.12
Consultant room 4.Waittime	2.22	15.13
Consultant room 5.Waittime	3.2	14.08
Xray machine 1.Waittime	3.2	8.29
Xray machine 2.Waittime	1.5	6.05

Table 14: Doctors' utilization result from OptQuest

Responses	**Value**	**Current state**
Doctor 1.Utilization	0.9	0.9752
Doctor 2.Utilization	0.85	0.9577
Doctor 3.Utilization	0.78	0.9066
Doctor 4.Utilization	0.7	0.8189
Doctor 5.Utilization	0.6	0.7721

5. Conclusion & Recommendation

5.1. Conclusion

This paper introduces an application of simulation on improving patients' waiting time at a private clinic. The discussion includes the research's scope, identify existing problem, modelling and propose a solution to help improve the efficiency of the clinic. With the use of simulation, the problem was easily identified that most of the congestion is around Operation Area, which is Consultant rooms and X-ray machines. Main problem was the inefficient appointment scheduling which causes the congestion in rush hours, by applying a tool in Arena simulation, OptQuest, the suitable arrival rate was determined to help improve the satisfaction of patients and the efficiency of the clinic. Further research could focus on applying new scheduling strategy, new optimize tool to produce even more improvement to the private clinic service.

5.2. Recommendation

Future research could expand the range of patients and clinics to other field or a general approach that apply similar clinics. In addition, a scheduling optimization for staffs or assign doctors to patients is another way to approach this problem.

Moreover, future study with more time the expand of data collection and more information can be collected to ensure 90-100% accuracy and produce a more optimize result. Another extension to this study is the appointment scheduling policy, a deep study with multiple policy can help strengthen the solution even more.

6. Acknowledgements

We would like to thank all my lecturers at the School of Industrial Engineering & Management of International University – Vietnam National University about all supports for the paper and Healthcare Clinic for Data Collection.

References

D. M. Ferrin, M. J. Miller, and D. L. (2007). McBroom, "Maximizing hospital finanacial impact and emergency department throughput with simulation," *Proc. - Winter Simul. Conf.*, no. January 2007, pp. 1566–1573, doi: 10.1109/WSC.2007.4419774.

J. K. Parks, P. Engblom, E. Hamrock, S. Satjapot, and S. Levin, (2011). "Designed to fail: How computer simulation can detect fundamental flaws in clinic flow," *J. Healthc. Manag.*, vol. 56, no. 2, pp. 135–144, 2011, doi: 10.1097/00115514-201103000-00009.

T. R. Rohleder, P. Lewkonia, D. P. Bischak, P. Duffy, and R. Hendijani, (2011). "Using simulation modeling to improve patient flow at an outpatient orthopedic clinic," *Health Care Manag. Sci.*, vol. 14, no. 2, pp. 135–145, 2011, doi: 10.1007/s10729-010-9145-4.

T. Cayirli and E. Veral, (2004). "Outpatient scheduling in health care: A review of literature," *Prod. Oper. Manag.*, vol. 12, no. 4, pp. 519–549, doi: 10.1111/j.1937-5956.2003.tb00218.x.

K. J. Klassen and R. Yoogalingam, (2009). "Improving performance in outpatient appointment services with a simulation optimization approach," *Prod. Oper. Manag.*, vol. 18, no. 4, pp. 447–458, doi: 10.1111/j.1937-5956.2009.01021.x.

B. Berg *et al.*, (2010). "A discrete event simulation model to evaluate operational performance of a colonoscopy suite," *Med. Decis. Mak.*, vol. 30, no. 3, pp. 380–387, doi: 10.1177/0272989X09345890.

C. Baril, V. Gascon, and S. Cartier, (2014). "Design and analysis of an outpatient orthopaedic clinic performance with discrete event simulation and design of experiments," *Comput. Ind. Eng.*, vol. 78, pp. 285–298, 2014, doi: 10.1016/j.cie.2014.05.006.

W. D. Kelton, (1983). *Simulation Analysis.*, vol. 1.

THE INTRINSIC ENTREPRENEURIAL MOTIVATION: A REVIEW ON THE DECISION OF STUDENTS TO PARTICIPATE IN THE PROJECT-BASED COMPETITION PROGRAM

Tran N.H. Ngan [1,3] and Andraz J.M.[1,2] and Nguyen H.L. Thuy [3]

[1] *Faculty of Economics, University of Algarve Campus de Gambelas, Building 9, 8005-139 Faro, Portugal, E-mail: ngan.tnh@vlu.edu.vn*

[2] *Faculty of Economics – University of Algarve and CEFAGE-UAlg - Center for Advanced Studies in Management and Economics, Campus de Gambelas, Building 9 8005-139 Faro, Portugal, E-mail: jandraz@ualg.pt*

[3] *Faculty of Commerce, Van Lang University, 68-69 Dang Thuy Tram, Ho Chi Minh City, Vietnam*

Abstract. This research explores on the deep mechanism that influences students in higher education institutions who voluntarily participate in the entrepreneurial project-based learning competition programs. We collect a survey from a convenience sample of 261 students from the business-related majors from different universities engaged in the Marketing Generators Project organized by the Faculty of Commerce, Van Lang University, Vietnam. We analyse our respondents' need for achievement, need for autonomy, calculated risks taking, drive and determination, and creative tendency in what concerns to their motivation to participate in the entrepreneurial competition for students. We suggest readers to view the results through the lens of self-determination theory and entrepreneurial behaviours. The findings indicate that need for achievement, calculated risks taking, and creative tendency are predictor of an individual's motivation to engage in different competitions and enrich personal knowledge and competence. Noticeably, the need for autonomy and independence as well as drive and determination has no explanatory power to form the intention of Vietnamese students in joining extra-curricular competitions. The findings could contribute with useful insights to understand students' studying motivation and to better design students-related projects which promote entrepreneurial behaviours of students at higher education institutions.

Keywords: entrepreneurial behaviour; project-based learning; intrinsic motivation; participation intention.

1. Introduction

Due to the past pace development of society, being competitiveness is the chance for business to survive and sustain (Newman, Tse, Schwarz, & Nielsen, 2018). This requires firms to have employees who not only are active, competitive and responsible but also engage in behaviours which are beyond their tasks and duties (Politis, 2005; Taşkıran & İYİGÜN, 2019). For that reason, the entrepreneurial behaviour and managerial attitudes toward risks of each individual working at the business have become valuable resources for the establishment of competitiveness of an organization (Taşkıran & İYİGÜN, 2019).

To prepare students with the professional knowledge and entrepreneurial skills required by the labour market, Higher Education Institutions (HEIs, here forth) have continually developed their programs to provide hard-skills and soft skills training for students (Toscher, 2019; Vogler et al., 2018). However, realizing that the traditional learning approach from which teachers take the role of "knowledge transmitters" and students act as "information receptors" is not sufficient in developing the readiness for students to fulfil the market requirement (Alorda, Suenaga, & Pons, 2011; Guo, Saab, Post, & Admiraal, 2020), project-based learning approach has been applied as it gives students opportunities to engage in a real business context with a real business problem from which their professional knowledge and transferable skills can be developed (Guo et al., 2020). Many studies have confirmed the positive results of project-based learning brought to students (Guo et al., 2020; Politis, 2005).

However, although there are many project-based learning programs organized by HEIs, the participation is not high despite the existence of a certain level of extrinsic rewards (Nyello, Kalufya, Rengua, Nsolezi, & Ngirwa, 2015). Deci and Ryan, in their Self-Determination Theory, argued that when people received extrinsic rewards for doing something, they are eventually less interested and engaged in it later (Deci & Ryan, 2008). The Organismic Integration Theory later supported this view by stating that if people perceive performing a task as their own choice leads to higher perceived autonomy and intrinsic motivation. This will make people value and engage more in the activity (Ryan & Deci, 2017). It is theoretically assumed that if people are high in self-determination, they will take higher responsibility for their actions to be competent and knowledgeable. This is when the learning outcomes can be achieved. Combined with entrepreneurial behaviour, this study examines the influences of the intrinsic motivation entrepreneurial motivation of students on their intention to participate in the project- based competition.

We offer the contribution to the understanding of the motivation to stimulate youngsters to actively challenge themselves through different professional projects and competition. This study is done following the self-determination theory and entrepreneurial behaviours. The survey uses samples of students who participated in the entrepreneurship workshop, organized as an orientation section for introducing the Marketing Generators project, from June to December 2020, organized by Van Lang University, Vietnam (n=261). The results indicate that need for achievement, calculated risks taking, and creative tendency are predictor of an individual's motivation to engage in different competitions and enrich personal knowledge and competence. However, the need for autonomy and independence and drive and determination have no explanatory power which can be explained by the fact that students in Vietnam are in their early stage of studying and working which still require more instruction from academic tutors and guidance from industrial experts to do real-life projects which they are not confident yet.

This article contributes to the existing research in three ways. We (1) provide further micro insights to HEIs of effective communication message design to encourage students to participate in project-based competition, as well as (2) a new application of entrepreneurial behaviours by applying it to study the intrinsic motivation of students in joining entrepreneurial programs. In doing so, we (3) expand and refine the psychometric survey items of this concept for the empirical setting of the entrepreneurial intention amongst students.

The rest of the paper is organized as follows. The next section describes the theoretical background and research hypotheses. Then, the paper discusses about data, variables and sample preliminary statistical analysis, econometric framework and estimation results. After that, the estimation of result and discussion are revealed. Finally, the last section concludes with some closing remarks.

2. Theoretical background and Research hypotheses

By definition, entrepreneurship, through their creation of new activities, promotes the development of a business or even an economy (Robles & Zárraga-Rodríguez, 2015). Entrepreneurial behaviour, is referred to the desire, motivation, competences and skills which individuals need for starting or managing a successful business (Robles & Zárraga-Rodríguez, 2015). As entrepreneurial skills and competences can be trained, numerous training programs have been developed and integrated into different levels with the aim of providing chances for individuals to acquire entrepreneurial skills and competences (Liñán, Rodríguez-Cohard, & Rueda-Cantuche, 2011). Further discussions of entrepreneurial behaviour theories, including the Self-determination theory and Intrinsic entrepreneurial motivation to project-based competition together with the research hypothesis will be elaborated.

2.1. Entrepreneurial behaviour Theory

Entrepreneurship behaviour and knowledge need to be nurtured amongst students as they aim to foster entrepreneurship mind-set, attitude, and skills needed to prepare graduates competence, behaviour, and readiness to be innovative in everyday business problem and to be proactive in utilizing resources and connecting with people to achieve business goals (Nyello et al., 2015; Politis, 2005; Toscher, 2019). Project based learning (PjBL) has been studied and applied by many HEIs with the focus on improving students' affective, cognitive, and behavioural outcomes which are needed to prepare them for employability (Guo et al., 2020). While affective outcomes reflect the understanding of students toward the benefits of participating in the projects, as well experiences accumulated from their involvement, cognitive outcomes are related to the formation of necessary knowledge and cognition to develop projects. Additionally, the behavioral outcomes refer to the extent the positive skills and behavior that emerge from the projects (Guo et al., 2020). Politis, in his research for entrepreneurial learning, indicates that entrepreneurial learning process is an experiential as knowledge needs to be acquired through the transformation in real cases (Politis, 2005). This does not minimize the needs of obtaining foundation knowledge and concepts obtained from the classroom. The inclusion of "discovery learning method", which allows students to learn how to solve problems and develop actions for a real case, is considered to offer the proper environment to challenge and guide entrepreneurial behaviour development (Abdelkarim, 2019).

Nyello et. al. (2015) has identified five essential entrepreneurial behaviours that motivate students in business majors to go beyond their normal training programs to develop continued entrepreneurial learning intention and practice outside their classroom context, including needs for achievement, needs for autonomy and independence, calculated risk taking, drive and determination, and creative tendency.

2.2. Self-determination Theory

Motivation gives people direction, persistence, energy, and intention to involve and activate an idea (Cnossen, Loots, & Witteloostuijn, 2019; Deci & Ryan, 2008). Self-determination theory (SDT, here forth) studies the people's profiles that can be self- motivated, self-regulated to take social functioning, and seek for personal fulfilment and well-being (Deci & Ryan, 2008). SDT consists of three basic psychological needs: needs for competence, needs for autonomy, and needs for relatedness (Deci & Ryan, 2008). First, needs for competence explains the degree of needs that people have to develop competence to master skills and tasks that are important to them. Second, needs for autonomy refers to the feeling of controlling their behaviour. Third, needs for relatedness concerns the sense of belongingness.

However, people can be affected by either self-motivation factors or external regulation factors (Deci & Ryan, 2008). Different from extrinsic motivation where motives to act rely heavily on external rewards, intrinsic motivation involves doing a behaviour because the activity itself is interesting, challenging, and rewarding (Deci & Ryan, 2008). Other authors argue that the overuse of extrinsic rewards would reduce intrinsic motivation across a range of age, rewards, and activities as the rewards themselves decrease the interest of people in doing the activity itself (Cnossen et al., 2019; Deci & Ryan, 2008). Research has shown that people with higher levels of need for competence and need for autonomy tend to be more innovative at work, more proactive in learning for mastering skills, value chances for development and success, and value the achievement of goals as rewards more than financial rewards (Abun et al., 2017; Bradley & Klein, 2016; Cnossen et al., 2019; Karabuluta, 2016; Karimi & Walter, 2016; Latif, Abdullah, & Jan, 2016; Mat, Maat, & Mohd, 2015; To, Martínez, Orero-Blat, & Chau, 2020)

Since the competition requires students from different universities to apply, to participate in intensive and short-time workshop training, and later to work in a group of five students for a specific local product which is appointed randomly by the organizer, needs for relatedness therefore cannot be formed. Deci and Ryan (2008) also highlighted that proximal relational support might not be needed for intrinsic motivation. For the reason, needs for relatedness is not considered in the research.

2.3. Intrinsic entrepreneurial motivation to project-based competition

Project-based learning (PjBL, here-forth) is the design of a learning activity where students are given an authentic business problem in a realistic context which is similar to what they will receive from the future workplace (Vogler et al., 2018). From participating in real-work setting projects, students not only have to find a way to work well with their team members but also to negotiate with other external stakeholders (namely contractor/ potential customers/ managers of companies) to understand the case, to plan and execute the project eventually. Hence, the project creates a learning and working environment from which students can develop both hard skills (understanding of the product, production process, consumers' behaviour for a specific product) and soft skills (communication skills, time management, and quality assurance of the project). With this real-life experience obtained from the workplace context, PjBL has proven to be a support to help students gradually to develop their cognitive knowledge, affective outcomes- understanding the importance of networking and activeness in participating in project-based learning, and behavioural outcomes which are skills of effective collaboration in the interdisciplinary team and other departments (Guo et al., 2020; Vogler et al., 2018).

Final competition activity between groups (CnBL, hereforth) aims at providing a healthy competition from which students, through their participation, can achieve their learning outcomes. However, the result from the competition is independent from students' academic scores (Alorda et al., 2011; Burguillo,2010; Cheng, Wu, Liao, & Chan, 2009). The combination of PjBL and CnBL is suggested to provide strong motivation for students to participate, while maintaining a healthy level of tension. Moreover, the winners of this competition are considered to receive social recognition as from other participants, organizers, schools, and companies.

2.4. The research hypotheses

This Project-based learning and competition between groups, namely The Marketing Generators, aimed to create opportunities for students belonging to business related majors of different undergraduate levels of different universities in Ho Chi Minh city to attend different training workshops before being introduced to different Small and Medium Enterprises, which are producing Sustainable Products in Vietnam to study about their cases and eventually to develop suitable Marketing plans for these firms.

Need for Achievement

Need for achievement is proven to be a predictor for entrepreneurial behaviour (Abun et al., 2017). Individuals with high need for achievement have higher desire for accomplishment and success which motivates them to set challenging goals, persistently pursue them and always seek for opportunities for mastering skills and knowledge (Gürol & Atsan, 2006; Karabuluta, 2016; Mat et al., 2015). With the inclination toward personal mastery and exploration, people are driven to join in different contextual events, to look for feedback, to look for cognitive and social development chances (Deci & Ryan, 2008). [Table 1] highlights the scale of the need for achievement. Hence, we expect that the potential feedback and competence and skills enhancement through participation in the competition and interaction with entrepreneurs and managers from local firms strongly motivate students to participate in the competition.

Hypothesis 1: *The Need for Achievement (NA) positively influences the intention to participate in the competition (IP)*

Need for Autonomy

Also, although intrinsic motivation gives people reasons, passion, and intention to work and achieve personal needs, theoretical and practical evidence show that the maintenance and enhancement of intrinsic motivation requires supportive environment and conditions (Deci & Ryan, 2008). The creation of an environment where people have choice, acknowledgement of feelings, and chances for self-direction, and which is created from voluntary project-based, is supposed to give people the feeling of autonomy (Deci & Ryan, 2008). [Table 2] highlights the scale of the for autonomy and independence. Thus, we expect that the sense of freedom to make decision on the marketing plan, the availability of deadline, pressure from the external evaluation, as well as the available guideline of academic mentor and industrial expert enables individuals to involve in the competition.

Hypothesis 2: *The Need for Autonomy and Independence (NAI) has a positive correlation with the intention to participate in the competition (IP)*

Calculated Risk Taking

Risk taking refers to the individual's tendency to take or avoid risk (S. N. Zhang, Li, Liu, & Ruan, 2020). Fear of failure often restraints people to proceed with a certain idea, a plan, or a course of action (Abun et al., 2017). In contrast, entrepreneurs are more comfortable with uncertainty and have higher tendency to make decision on incomplete information, as well as to take calculated risks taking base on good evaluation of the likely benefits against the likely costs of actions (Abun et al., 2017; Karabuluta, 2016; S. N. Zhang et al., 2020). [Table 3] highlights the scale of the need for risk taking. Hence, calculated risk taking is considered to be a potential predictor of the intention to participate in the competition which is considered as a new program and not compulsory to studying programs which might take a lot of time and effort apart from students' tight studying schedule. Not to mention, this competition also requires students to work in teams involving students from different universities and majors.

Hypothesis 3: *The Calculated Risk Taking (CRT) has positively increased the intention to participate in the competition (IP)*

Drive and Determination
Drive and determination refer to the inner motivation that drives people to get out of their comfort zone and look for new opportunities to pursue a certain activity or business (Abun et al., 2017). Drive refers to a psychological need which encourages individuals to go into actions and persistently work to achieve success on their terms (Caird, 2013). In the context of personal development, not only drive refers to motivation but also reflects the confidence and the ability to take action and responsibility for their learning (Abun et al., 2017). Additionally, determination inspires people to continuously overcome arising obstacles to achieve their goals. [Table 4] highlights the scale of drive and determination. Powered by drive and determination, students are believed to take action, to participate and persistently work hard to the final round in the competition.

Hypothesis 4: *The Drive and Determination (DD) is positively correlated with the intention to participate in the competition (IP)*

Creative Tendency
Creativity reflects the ability to evaluate the market, to identify new opportunities, and the tendency to generate ideas and alternatives to solve certain business problems (Abun et al., 2017; Newman et al., 2018; S. N. Zhang et al., 2020). Indeed, creativity has been proved to have influence not only on entrepreneurs but also on employees' entrepreneurial behavior and intention to be more creative at work (Newman et al., 2018). With strong passion on being innovative, individuals often opt for change-oriented, resulting in the activeness to involve in the process of learning, identifying problems and drawing personal resources to generate ideas to solve problems (Caird, 2013; Gürol & Atsan, 2006; Karimi & Walter, 2016;
S. N. Zhang et al., 2020). [Table 5] highlights the scale of creative tendency. For that reason, we expect that students with creative tendency, which have exploration spirit and creative thinking, will actively look for chances to be involved in the creative process and generating ideas to help local business.

Hypothesis 5: *The Creative Tendency (CT) positively influences the intention to participate in the competition (IP).*

3. Data, variables and sample preliminary statistical analysis

3.1. The institution context and sample
This research uses data collected from an online survey sending to students who studied business and the ones participated in the Marketing Generation competition organised by a private university in Vietnam. Like other private institutions, it faces challenges in diversifying and increasing its academic values offered to students to consolidate its ranking amongst students and society. Developing workshop training, field-trip to different companies and countries, as well as applying project based learning and competition are few practices have been conducted by most private universities. Unfortunately, low participation rate of students to these activities despite effort of institutions in organizing such activities demands deeper understanding of what motivates students to participate in extra-curricular training programs.

This paper seeks to examine the influences of entrepreneurial intrinsic needs to the intention to participate in project based learning and competition by using quantitative data analysis. A survey was administered and delivered to students who participated in the entrepreneurship workshop, organized as an orientation section for introducing the Marketing Generators project, from June to December 2020. A sample size of 261, accounted for 62 percent of the participants of the workshop is satisfactory to carry the research with the number of variables in our study (Cnossen et al., 2019).

3.2. Variables
The research instrument was structured into three parts. The first part included socioeconomic variables. The second part consists of five variables to measure need for achievement, need for autonomy and independence, calculated risk taking, drive and determination, and creative tendency which are believed to have intrinsic motivation effects to the intention of students to participate in the competition project. For this part, respondents were asked to indicate the agreement or disagreement with each statement given on a five-point Likert scale. The third part measures the intention of students whether they will participate in the project.
Each variable in the second part was measured by five items synthesized from studies of several authors. For all the scales, responses were coded in a way that a higher score reflects a higher need for achievement, need for autonomy and independence, calculated risk taking, drive and determination, and creative tendency.

3.3. Data preliminary statistical analysis: Descriptive statistics and Principal Components Analysis
Socioeconomic analysis has been described in term of four areas, including gender, year of study, studying major, and occupation of parents. Table 1 illustrates the frequency of the analysis. In term of gender, most of respondents are female, which was accounted for

76.6% and male respondents accounted for 23.4%. This shows that female respondents are more responsive to the survey than male. In term of year of study, most of respondents are freshman (49.4%), followed by students in year 3 (22.2%) and sophomore (21.8%). The least number of respondent belonged to students of the final year, measured at 6.5%.

Further, in term of studying major, the dominant respondents belonged to marketing major (40.2%) and followed by economics and commerce (34.1%), business administration (16.9%) and other majors (8.8%). As regard to parents' occupation, it seems that parents with professional jobs and hold managerial position have stronger connection with students who responded to the survey, accounting for 20.7% and 23.4% respectively.

Cronbach's Alpha is used to measure internal consistency of the construct and test how well the items are correlated to each other. A scale is regarded as reliable when the Cronbach's Alpha is around and even higher than 0.7. As all constructs have the Cronbach's Alpha higher than 0.869, all constructs are considered as having good internal consistency.

The 25 items were subjected to a principle component analysis (EFA) with KMO and Barlett's test and Varimax normalized rotation. KMO (0.952) and Sig. (0.000) confirm that the null hypothesis is rejected and the existence of relationship between variables.

Estimated communalities exhibit high values. The total explained variance is above 63% and the variation suggests the extraction of 3 factors, which are built on the dimensions with the highest number of items' loadings above 0.60. After performing rotated component matrix, only the remaining scale items of five variables which have the high loading factors (higher than 0.6) are considered for modelling purposes.

4. Econometric framework and Estimation results

Binary logistic regression model has been used to analyse the relationship between independent and dependent variables, given the nominal nature of dependent variable which is latent and cannot be observed.

This setting allows us to measure the likelihood of each input variable on the intention to participate to the competition; thus, answering the research questions. A binary regression is first conducted to eliminate statistically non-significant explanatory variables with p-value is lower than 0.05 and Wald statistic is smaller and 4. The result of this regression suggests the elimination of few variables belonging to each factor. For the model specification purpose, the chosen satisfaction predictors which are in bold as followed:

- NA_4 – Need for achievement
- CRT_3 - Willingness to seek for help
- CT_1 - Creative tendency in making solutions

The model for estimating the intention to participate in the entrepreneurial competition is formulated as following:

$$ln\frac{P(\gamma=0)}{P(\gamma=1)} = \beta_0 + \beta_1(NA_4) + \beta_2(CTR_3) + \beta_3(CT_1) + \mu \quad (1)$$

where γ repreents the dependent variable "Intention";

β_j (for j = 0; 1; 2; 3) is the equations' parameter; μ is the usual stochastic error term;

$ln\frac{P(\gamma=0)}{P(\gamma=1)}$ represents the logarithm of the odds ratio of intention to participate in the competition.

Using some algebraic manipulation, the expressions for the estimated probabilities associated to the "Participation intention" variable, given the values of the covariates of no intention and with intention are as follows:

$$P(\gamma = 0|NA, CRT, DD, CT) = \frac{e^{ln\frac{P(\gamma=0)}{P(\gamma=1)}}}{1+e^{ln\frac{P(\gamma=0)}{P(\gamma=1)}}} \quad (2)$$

5. Estimation result and Discussion

This section presents the estimation results and the interpretations of the results, including the coefficients' estimates of the model and the relative importance of each variable to the participation intention of students.

5.1. Coefficient estimates and diagnostic testing

The estimation results for the equation, reflects useful information insights on the importance of each covariate on the logs of the odds ratios in each model and therefore on the relative likelihood of participation intention.

The estimated coefficients are significant with p-value lower than 0.05 and Wald statistic is higher or lower than 4. Moreover, all coefficients are positive. The parameter estimates associated to the NA_4 – Need for achievement, CRT_3 - Willingness to seek for help, CT_1 - Creative tendency in making solutions are 1.019, 1.147, and 1.076 respectively. These values represent the increase of 2.770, 3.149, and 2.933 on the "oddratio" (note that $e^{1.019} = 2.770$; $e^{1.147} = 3.149$; $e^{1.076} = 2.933$) leads to an increase in P for intention to participate in the programs.

5.2. Discussion

Hypothesis 1 has been confirmed: The Need for Achievement (NA) positively influences the intention to participate in the competition.

Hypothesis 3 has been confirmed: The Calculated Risk Taking (CRT) has positively increased the intention to participate in the competition.

Hypothesis 5 has been confirmed: The Creative tendency (DD) is positively correlated with the intention to participate in the competition.

However, the variable "Need for autonomy and independence" (NAI) has no explanatory power and **hypothesis 2** needs to be rejected. This rejection can be explained based on Hofstede's cultural dimension theory (Hofstede, 2001). Vietnamese people perceived fairly high collectivism, as the score of individualism dimension quite low (M. Nguyen & Truong, 2016; Tran, 2009). This explains why students might not be ready to work independently and have higher fear of working with strangers in an irregular competition setting and become reluctant to apply for the project-based learning and competition.

Also, **hypothesis 4** is also rejected as The Drive and Determination (DD) has no explanatory power to the intention to participate in the competition. based on Hofstede's cultural dimension theory, there is a large power distance in Vietnamese culture (Hofstede, Hofstede, & Minkov, 2010; Truong, Hallinger, & Sanga, 2017). Together with high collectivism, this trait also explains why students, who are in their early stage of studying and working, still require more instruction from academic tutors and guidance from industrial experts to do real-life projects which they are not confident yet (H. T. M. Nguyen, 2016). However, since they are reluctant to work closely with tutor, there is a higher chance of either not participating in the competition or withdrawing from the competition after attending the training workshop.

6. Conclusion

This study intends to identify the determinants to participate in entrepreneurial competition of students in business related majors in HEIs by analysing a particular case of Van Lang University, one Vietnamese private university. From the literature review, needs for achievement, need for autonomy and independence, calculated risk taking, drive and determination, and creative tendency are all intrinsic motivations that drive students to join in extra-curricular activities to challenge themselves and improve their personal knowledge, skills, and experience.

6.1. Main conclusion

In general, students with high need for achievement who values accomplishment and achievement, even more than financial rewards, are more likely to challenge themselves in registering for the entrepreneurial competition. Similarly, those are willing to seek for help, even from strangers, demonstrates a stronger willingness to take calculated risk and intention to participate in these project-based competitions. Lastly, students who have the creative tendency and always look for new alternative solutions for a given problem also demonstrates higher intention to participate the competition.

The results suggest that the resistance to join in different extra-curricular is the main cause of low participation rate of students to different project-based learning and competitions organized by HEIs. Thus, apart from the effort put in organizing the extra- curricular activities, HEIs should initially focus on giving orientation to highlight the gains obtained from these programs to students' knowledge, skills, and employability possibility. This might motivate students to improve the need for achievement and actively look for more opportunities to accumulate achievements and experiences. Secondly, as the projects are new to students, making process supports and academic guidance available might help reduce the risk perceived by students and eventually convince them to join the competition. Thirdly, more soft skills training workshops relating to team-working, project management, creative thinking, critical thinking and other related project training are needed to provide foundation knowledge and means for students to know how to generate innovative ideas and take further actions.

6.2. Contributions to entrepreneurship education research

(Karabuluta, 2016) and (Abun et al., 2017) have identified the influences of entrepreneurial attitude on intentions while (Cnossen et al., 2019) have proven that creative entrepreneurship is more intrinsically than extrinsically motivated. In line with existing research, this study hence provides a further confirmation on the effects of intrinsic entrepreneurial attitude as determinants which motivate students to develop intention to participate in entrepreneurial training and competition projects. Moreover, the estimation of the binary logistic regression model allows the identification of the predictive intention of each variable construct to the intention of students. The findings also reveal that students with different cultural background will be motivated differently by different attributes. In this case, need for autonomy and independence has minor effect.

6.3. Management implications

Our findings show information about three aspects which should be considered as determinants to the participation of students to the project-related training programs, including need for achievement, calculated risk taking, and creative tendency. These determinants are closely connected with intrinsic motivation which gives Vietnamese students reasons and motives to pay attention to the programs, to do further research and later make decision to join in these programs. Therefore, as entrepreneurship can be learned and trained and as there is the willingness to learn from students, these entrepreneurial competition projects are necessary to initially develop entrepreneurial attitudes and confidence which are needed for students in their later career (Abun et al., 2017; Guo et al., 2020; Gürol & Atsan, 2006).

6.4. Limitations and suggestions for further research

Our paper focuses on a case study from which the findings might not be generalised and applicable to all business-majored students in Vietnam and other countries. Moreover, as this study only focuses on identifying the influences of intrinsic motivation to the intention of participating in entrepreneurial project-related training programs, more studies should be done to compare the influences of both extrinsic and intrinsic motivations on students' intention. Additionally, apart from assessing the influences of motivation factors, the identification of other resistance factors which constraint students from attending these programs should also be identified. Moreover, the outcomes of such participation in term of knowledge, skills, attitude gained from these project-related training programs should be measured to evaluate their effectiveness. Finally, the small sample size does not allow us to draw the general causality and it does inspire future research with quantitative data collection on a larger scale.

Reference

Abdelkarim, A. (2019). Toward establishing Entrepreneurship Education and Training Programmes in Multinational Arab University. *Journal of Education and Training Studies, 7*(1), 1-9.

Abun, D., Lalaine, S., Foronda, G., Agoot, F., Luisita, M., Belandres, V., & Magallanez, T. (2017). Measuring entrepreneurial attitude and entrepreneurial intention of ABM grade XII, Senior High School Students of Divine Word Colleges in Region I, Philippines. *Academy of Natural Sciences, 4*(4), 100-114.

Alorda, B., Suenaga, K., & Pons, P. (2011). Design and evaluation of a microprocessor course combining three cooperative methods: SDLA, PjBL and CnBL. *Computers & Education, 57*, 1876–1884.

Bradley, S. W., & Klein, P. (2016). Institutions, economic freedom, and entrepreneurship: The contribution of management scholarship. *The Academy of Management Perspectives, 30*(3), 211-221.

Burguillo, J. C. (2010). Using game theory and Competition-based Learning to stimulate student motivation and performance. *Computers & Education, 55*, 566-575.

Caird, S. (2013). General measure of Enterprising Tendency test (GET). Retrieved from https://www.researchgate.net/publication/301347695_General_measure_of_Enterprising_Tendency_testwwwget2testnet

Cheng, H. N., Wu, W., Liao, C. C., & Chan, T. W. (2009). Equal opportunity tactic: redesigning and applying competition games in classrooms.

Computers & Education, 53(3), 866-876.

Cnossen, B., Loots, E., & Witteloostuijn, A. V.

(2019). Individual motivation among entrepreneurs in the creative and cultural industries: A self-determination perspective. *Creativity and Innovation Management*, 1-14.

Deci, E. L., & Ryan, R. M. (2008). Facilitating Optimal Motivation and Psychological Well- Being Across Life's Domains. *Canadian Psychology, 49*(1), 14–23.

Douglas, E. J., & Shepherd, D. A. (2002). Self- employment as a Career Choice: Attitudes, Entrepreneurial Intentions, and Utility Maximization. *Entrepreneurial Theory and Practice, 26*(3), 81-90.

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research, 102*, 1-13.

Girol, Y., & Atsan, N. (2006). Entrepreneurial characteristics amongst university students. *Education and Training, 48*(1), 25 - 38.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. London: Sage publications.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010).

Cultures and organizations: Software of the mind - Intercultural cooperation and its importance for survival (3rd edition ed.). New York.: McGraw Hill

Karabuluta, A. T. (2016). Personality Traits on Entrepreneurial Intention. *Procedia - Social and Behavioral Sciences, 229*, 12 – 21.

Karimi, J., & Walter, Z. (2016). Corporate Entrepreneurship, Disruptive Business Model Innovation Adoption, and Its Performance: The Case of the Newspaper Industry. *Long Range Planning, 49*, 342–360.

Latif, N. S. A., Abdullah, A., & Jan, N. M. (2016). A Pilot Study of Entrepreneurial Orientation towards Commercialization of University Research Products. *Procedia Economics and Finance, 37*, 93 – 99.

Liñán, F., Rodríguez-Cohard, J. C., & Rueda-Cantuche, J. M. (2011). Factors affecting entrepreneurial intention levels: a role for education. *International Entrepreneurship and Management Journal, 7*, 195–218.

Mat, S. C., Maat, S. M., & Mohd, N. (2015).
Identifying Factors that Affecting the Entrepreneurial Intention among Engineering Technology Students. *Procedia - Social and Behavioral Sciences, 211*, 1016 – 1022.

McMullen, J. S., Bagby, D. R., & Palich, L. E. (2008). Economic Freedom and the Motivation to Engage in Entrepreneurial Action. *Entrepreneurship Theory and Practice*, 875-895.

Newman, A., Tse, H. H. M., Schwarz, G., & Nielsen,
I. (2018). The effects of employees' creative self-efficacy on innovative behavior: The role of entrepreneurial leadership. *Journal of Business Research, 89*, 1-9.

Nguyen, H. T. M. (2016). The influences of 'power distance'on pre-service teacher learning in Vietnam. . *Journal of Nusantara Studies (JONUS), 1*(2), 38-49.

Nguyen, M., & Truong, M. (2016). The Effect of Culture on Enterprise's Perception of Corporate Social Responsibility: The Case of Vietnam. . *Procedia CIRP, 40*, 680-686.

Nyello, R., Kalufya, N., Rengua, C., Nsolezi, M. J., & Ngirwa, C. (2015). Effect of Entrepreneurship Education on the Entrepreneurial Behaviour: The Case of Graduates in the Higher Learning Institutions in Tanzania. *Asian Journal of Business Management, 7*(2), 37-42.

Politis, D. (2005). The Process of Entrepreneurial Learning: A Conceptual Framework.
Entrepreneurship Theory and Practice, 399-424.

Robles, L., & Zárraga-Rodríguez, M. (2015). Key Competencies for Entrepreneurship. *Procedia Economics and Finance, 23*, 828 – 832.

Ryan, R. M., & Deci, E. L. (2017). *Self- deteremination theory: Basic psychological needs in motivation, development, and wellness*. New York: The Guilford Press.

Taşkıran, G., & İYİGÜN, N. Õ. (2019). The
Relationship between Organizational Citizenship Behavior and Entrepreneurial Orientation: A Research in the Hospitality Industry. *Procedia Computer Science, 158*, 672–679.

To, C. K. M., Martínez, J. M. G., Orero-Blat, M., & Chau, K. P. (2020). Predicting motivational outcomes in social entrepreneurship: Roles of entrepreneurial self-efficacy and situational fit. *Journal of Business Research, 121*, 209–222.

Toscher, B. (2019). Entrepreneurial Learning in Arts Entrepreneurship Education: A Conceptual Framework. *A Journal of Entrepreneurship in the Arts, 8*(1), 3-22.

Tran, H. (2009). *The Transformation of Higher Education in Vietnam after DoiMoi: A Story of "Dualism".* (Dissertations/Theses - Doctoral Dissertations). St. John's University, New York. (ED526932)

Truong, T. D., Hallinger, P., & Sanga, K. (2017).
Confucian values and school leadership in Vietnam: Exploring the influence of culture on principal decision making. . *Educational Management Administration & Leadership, 45*(1), 77-100.

Vogler, J. S., Thompson, P., Davis, D. W., Mayfield,
B. E., Finley, P. M., & Yasseri, D. (2018). The hard work of soft skills: augmenting the projectbased learning experience with interdisciplinary teamwork. *Instructional Science, 46*, 457–488.

Zhang, P., Wang, D. D., & Owen, C. L. (2015). A
Study of Entrepreneurial Intention of University Students. *Entrepreneurship Research Journal, 5*(1), 61-82.

Zhang, S. N., Li, Y. Q., Liu, C. H., & Ruan, W. Q.
(2020). Critical factors identification and prediction of tourism and hospitality students' entrepreneurial intention. *Journal of Hospitality, Leisure, Sport & Tourism Education, 26*, 1-18.

Appendix.

Table 1. Scale for the construct: Need for achievement

Subtitles	Designation	References
Desire for accomplishment	Desire for significant accomplishment and success	(Karabuluta, 2016; Mat et al., 2015)
Desire for mastering of skills	Desire for mastering of skills, control, or high standards	(Karabuluta, 2016; Mat et al., 2015)
Desire for challenging goal	Setting challenging but realistic goals	(Gürol & Atsan, 2006; Karabuluta, 2016; Mat et al., 2015)
Value achievement	Value the achievement of goals as rewards more than financial rewards	(Karabuluta, 2016; Mat et al., 2015)
Result oriented	An interest in observing results of their decision	(Abun et al., 2017; Karabuluta, 2016)

Table 2. Scale for the construct: Need for autonomy and independence

Subtitles	Designation	References
Freedom to express idea	Freedom to express personal views	(Bradley & Klein, 2016; McMullen, Bagby, & Palich, 2008; To et al., 2020)
Independence	Independence – being able to exercise the plan independently	(Bradley & Klein, 2016; Douglas & Shepherd, 2002; Gürol & Atsan, 2006; McMullen et al., 2008; To et al., 2020)
Strong self-expression	Strong self-expression, feeling the need to do what they want to do in their way, rather than work on other people's initiatives	(Bradley & Klein, 2016; Karimi & Walter, 2016; McMullen et al., 2008; To et al., 2020)
Willingness to defend personal idea	Unconventional, being prepared to stand out as being different to others and defend for personal idea	(Abun et al., 2017)
Belief in multitasking ability	Belied in personal ability to perform effectively on many different tasks	(Bradley & Klein, 2016; Douglas & Shepherd, 2002; Gürol & Atsan, 2006; McMullen et al., 2008; To et al., 2020)

Table 3. Scale for the construct: Calculated risk taking

Subtitles	Designation	References
Make decision in uncertainty	Make decision in an uncertain situation/ Tolerate risks & uncertainty more than others	(Abun et al., 2017; Douglas & Shepherd, 2002; Gürol & Atsan, 2006; Karabuluta, 2016; Karimi & Walter, 2016; P. Zhang, Wang, & Owen, 2015)
Risk taking	Categorized risky situations as positive	(Karabuluta, 2016)
Willing to seek for help	Actively ask for help and advice from other people	(Abun et al., 2017)
Calculated risk taking	Analytical, being good at evaluating the likely benefits against the likely costs of actions	(Abun et al., 2017; Karabuluta, 2016; S. N. Zhang et al., 2020)
Taking responsibility	Decisive responsible for actions	(Gürol & Atsan, 2006; Karimi & Walter, 2016)

Table 4. Scale for the construct: Drive and determination

Subtitles	Designation	References
Belief in personal ability	Self-confidence with the belief that they have control over their destiny and make their own luck, rather than being controlled by fate	(Abun et al., 2017; Caird, 2013; Gürol & Atsan, 2006)
Drive	Outcome oriented, always complete what has been planned	(Abun et al., 2017; Caird, 2013; Latif et al., 2016)
Turn problems into opportunities	Ability to turn problems into opportunities	(Abun et al., 2017; Gürol & Atsan, 2006; Latif et al., 2016)
Determination	Proactive, taking personal responsibility to navigate the problems that arise to achieve success on their terms	(Caird, 2013)
Opportunistic	Opportunistic, seeking and taking advantage of opportunities	(Abun et al., 2017; Gürol & Atsan, 2006)

Table 5. Scale for the construct: Creative tendency

Subtitles	Designation	References
Creative tendency in making solutions	Imagination, with an inventive or innovative tendency to come up with new ideas	(Abun et al., 2017; Caird, 2013; Karimi & Walter, 2016; Latif et al., 2016; Newman et al., 2018; S. N. Zhang et al., 2020)
Change-orientated	Change-orientated, preferring novelty, change and challenge	(Abun et al., 2017; Gürol & Atsan, 2006; Karimi & Walter, 2016; Newman et al., 2018)
Encourage others to be innovative	tendency to encourage others to generate more possibilities.	(Abun et al., 2017; Gürol & Atsan, 2006; Karimi & Walter, 2016; Newman et al., 2018)
Willing to deal with obstacles	Being able to draw on personal resources for projects or problem solving	(Abun et al., 2017; Caird, 2013; Karimi & Walter, 2016)
Realistic creation	Tendency to have a realistic innovation, devote a great deal of time to collect information to find solutions for real-work issues.	(Abun et al., 2017; Caird, 2013; Karimi & Walter, 2016)

AUTOMATIC PROCESS RESUME IN TALENT POOL BY APPLYING NATURAL LANGUAGE PROCESSING

Quynh Trinh[1*], Thanh-Tuan Dang[1]

[1] Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

**Corresponding author: f108143135@nkust.edu.tw (Q.T)*

Abstract. The fast-paced development of technology gradually solves the challenges of resource and time-consuming in recruitment. For the last few years, going hand-in-hand with natural language processing growth, this area has revolutionized the quality of insights and partly automated to accelerate manual tasks. In this paper, a new framework to efficiently build a talent pool with qualified and appropriate candidates is proposed as an effective solution to improve recruitment quality and cost-saving. The talent pool is regarded as a database where recruiters and Human resource (HR) departments keep all of the candidates who have applied for jobs and are sourced, referred, qualified, or interested. Moreover, this kind of database is common in headhunter company to manage a huge number of resumes and job orders. In today's world of data-driven recruiting, establishing a talent pool has become extremely important, especially in head-hunter or large companies. The proposed framework is divided into two phases. Phase 1: the proposed method to extract features and topics from a large number of e-resumes and job posts. Phase 2: our framework classifies and ranks e-resumes according to occupational categories based on the features obtained from phase 1. A strong baseline is constructed to fulfil the critical gap in pooling data. This paper's main contributions are two-fold: (1) we propose a theoretical framework to enhance the talent pool construction procedure, making the recruitment procedure more efficient and accurate; (2) we present a new approach in ranking resumes based on practical heuristics. This proposed framework can be a valuable reference for decision-makers in automatically processing resumes or developing relevant systems.

Keywords: Human resource management, natural language processing, talent pool, cosine similarity

1. Introduction

Along with technology development, Human Resource Management (HRM) core functions currently exploit and apply technology to reduce the manual workload and achieve better management and time-saving. In the interim, online application and recruitment are more common and become unavoidable in the digital HR transformation age. However, due to the high flow of new candidates, enhancing database management is a big gap needing fulfillment to avoid data overwhelms and time-consuming for recruiters. This paper proposed a feasible framework to approach this problem by enhancing the database known as the company's talent pool.

When having a new job vacancy, sourcing new e-resumes and databases is time-consuming and manual workload. Furthermore, this sourcing activity can be affected by personal bias and specific knowledge limitations. The HR department has to face various occupational categories such as hardware, software, firmware, marketing, sales, and general back office. To improve the recruitment quality, the recruiters are also required to have a certain knowledge of the occupational categories they are dealing with. Typically, each company and recruiter have their database of e-resumes which is known as the talent pool. To be more specific, a talent pool is the database-driven administration of profiles from the company's employees to various sources from career fairs, social media, and employee referrals, to name a few.

It could be merely internal or personal management, an online platform, or accessed through the company website's careers page. A talent pool is essential for successful business management, especially by enabling efficient talent acquisition, optimizing active sourcing, and targeting HR development. In a more extensive overview of strategic workforce planning, long-term recruitment should include a well-maintained talent pool in their strategy. In the initiative support, developed software solutions such as Bullhorn and SAP SuccessFactors enable recruiters to manage professional profiles and access the relevant information. However, through the availability of new digital technologies of artificial intelligence, there has been a new HRM transformation that is undertaking by Natural Language Processing (NLP) and Machine learning techniques.

In pure expressions, NLP represents the automatic approach to natural human languages under speech or text formation. NLP has been widely used in many areas such as email classification, fake news detection, healthcare, cognitive assistant in electronic devices, information identification, and extraction for business purposes. Velay et al. (2018) used NLP to analyze new headlines to forecast the DJIA (Dow Jones Industrial Average) trend. With the same method, Yıldırım et al. (2018) also worked on the headline to detect the financial trend. In fake news detection, Kong et al. (2020) and Bourgonje et al. (2017) developed methods based on NLP to examine online articles' headline and content. In healthcare concerns, Kaufman et al. (2016) and Byrd (2014) performed NLP on the process of documentation in electronic health records (EHRs) to the better treatment regimen. In the interim, the processing of natural language is also used in Human Resource Management. Jung et al. (2019) applied Latent Dirichlet Allocation (LDA) to identify job satisfaction factors from 35,063 online employee reviews posted on jobplanet.co.kr. Menon et al. (2016) presented an approach to evaluate and rank candidates based on estimating emotional intelligence through their medial social data. Previous studies have shown the feasibility of utilizing NLP in the recruitment procedure. Amin et al. (2019) developed the web application to scan resumes for a particular job posting. The machine learning and NLP techniques were used to compare resumes with job posts. However, it is an interactive website that needs job applicants to verify their information. Gopalakrishna et al. (2019) focused on classifying resumes in the IT sector. To automatic resumes processing for scientific research, El Mohadab et al. (2018) used a data mining algorithm. Za-roor et al. (2017) built a system to classify the resumes and match them with the job posts. For the same purpose, Bhatia et al. (2019) proposed a framework using the BERT technique, but this paper restricts the candidate's past professional experience. However, no previous research has been focused on studying enhancing the talent pool by apply NLP techniques.

Regarded to other techniques used in this paper, cosine similarity, a method to measure how similar the documents are based on cosine of the angle between two vectors projected in a multi-dimensional space. The higher the cosine similarity, the smaller the angle is. In NLP, this method is widely applied. In this paper, we utilized this method to compute the similarity between job post and resume to generate the suitable recommendation list. As mentioned, the talent pool plays an essential role in the recruitment function in human resource management. Despite its necessary and resourceful data, the talent pool is often seen as the second choice when recruiters search for an opening job. The time of maintenance, adding a new database, and limited searching algorithms in existing pools consume a significant amount of time. This paper proposes a framework to enhance existing talent pool systems by employing NLP techniques and word embedding. The framework automatically handles the input resumes by classifying and ranking them in their appropriate predefined occupational categories. Each category contains a list of belonging job titles and their descriptions accordingly. The expected output is the list of resumes, which sort the job-matching grades in each occupational category's descending order. Therefore, when there is a job vacancy, the list of top candidates is available for further steps. Overall process is shown in Fig. 1

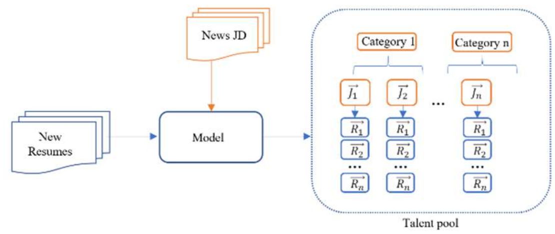


Figure 1. Talent pool

In summarization, this paper aims to fill the critical gap of establishing a talent pool regarding human resource management's scope under long-term strategic workforce planning. This paper's contributions are: (1) propose a framework to enhance the talent pool through optimizing the staffing workload, time-saving and efficient talent acquisition; (2) present a new approach in resume classification and ranking. This proposed framework can be a valuable reference for related decision-makers in automatic process resume or developing relevant systems.

2. Framework Proposal

Practically, the HR department receives resumes through various and flexible sources, which can be named a few as cumulative applications to opening jobs, career fairs, referrals, current, former employees, and recruiters' active search. These resumes are stored in the company database, mentioned as the talent pool, in various formats. The proposed framework's first step is to pre-process the text of resumes and job descriptions and perform personal information extraction. Through this step, the expected output is processed resumes and job descriptions, and retrieved information of general personal information (gender, year of birth, address), education (the highest degree, certificate), and languages in the resumes. In the second step, the text-processed resumes and job descriptions are embedded and classified into occupational information and general information. The outputs are used to present the resume and job description vectors and employed as the input for similarity measurement between resume and job description. Finally, to improve the talent pool's quality, the weights of personal information extraction in the first step are added to compute the final scores.

The proposed framework's output is the resume list, which sorts the grades in descending order in each occupational category. Therefore, when a job opens, the list of top candidates is available for the recruiter to proceed. The resume and job description (JD) have proceeded the same process in this proposed framework yet separately.

The proposed framework has three main parts, shown in Fig. 2.

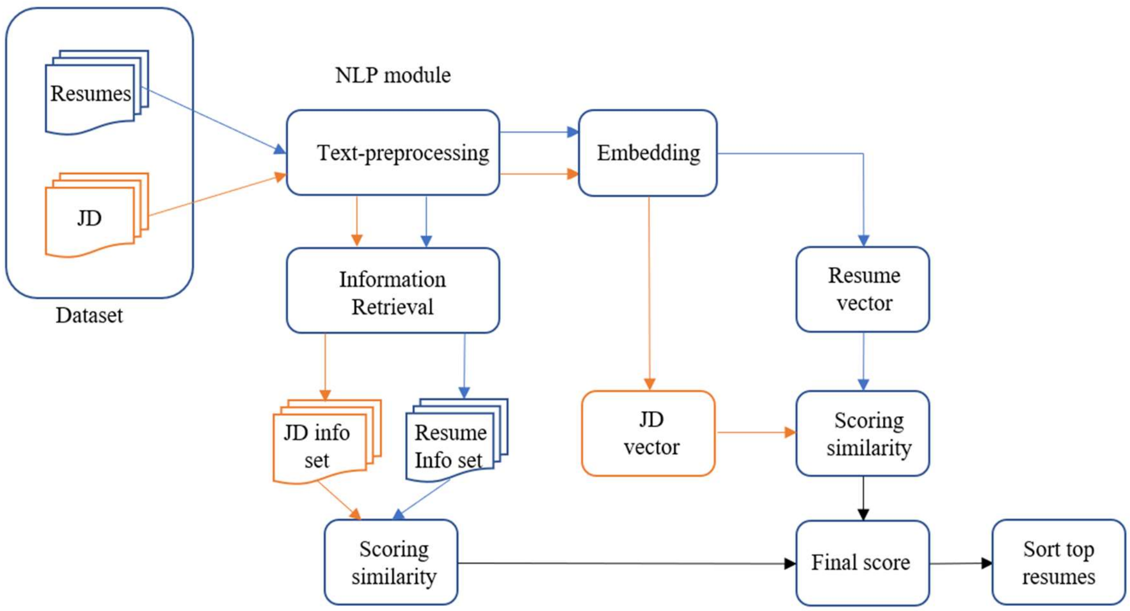


Figure 2. Overall framework

2.1. Preprocessing data and information

Aforementioned, the number of resumes in a talent pool is vast and comes from various channels. The writing styles are also diversified. Therefore, in this part, the unnecessary information, mistypes, and duplication are removed. After the text pre-processing steps, we perform parsing with custom NER training with SpaCy to extract the personal data. As the characteristic of recruitment, the candidates' gender, address, degree, and languages are affected by the possibility of job matching. These features' information is independently extracted, and their results are stored in a separate set for pair matching with those from JD.

2.2. Resume and JD's embedding and segmentation

In this second step, the text-processed resume and JD are embedded into the linear space Rn. The word embedding method applied here is the "fastText" which is an extension of the Word2vec model. This embedding method represents each word as an n-gram of characters. Thus, it can provide vector representation for the words which are not in the model dictionary. New jobs and terms are gradually added in due to the fast pace of change and personal writing style. Therefore, this method is expected to come across this problem.

After performing embedding, the problem of segmentation is regarded as two binary classification tasks. The resume and JD are segmented into two main parts, which are occupational information and general information. To calculate the similarity between resume and JD, we represent them as a space vector. For example, the job description is expressed as a vector contained (occupational job information, general job information). It is represented by $\vec{J} = (\vec{J_1}, \vec{J_2})$; student resume is expressed as a vector like this (occupational resume information, general resume information). It is represented by $\vec{R} = (\overrightarrow{R_1}, \overrightarrow{R_2})$. The similarity between two jobs or two students can be calculated by the formula (1) and (2):

$$\text{simi}\left(\overrightarrow{J_1}, \overrightarrow{J_2}\right) = Cos\theta_j \quad (1)$$

$$\text{simi}\left(\overrightarrow{R_1}, \overrightarrow{R_2}\right) = Cos\theta_u \quad (2)$$

Set and parameter of formulations:

$simi\ (\overrightarrow{J_1}, \overrightarrow{J_2})$	similarity between JD_1 and JD_2
$simi\ (\overrightarrow{R_1}, \overrightarrow{R_2})$	similarity between R_1 and R_2
$cos\ (\theta)$	cosine of the angle created between vectors

2.3. Resume and JD's embedding and segmentation

Due to the characteristics of specific parts in a resume, the weights of personal information extraction in the first step are added to computing the final scores to improve the talent pool's quality.

In the real world, some job descriptions can require specific factors from gender, address, degree, and languages such as gender limited to male or female only. We calculate each factor and finally add in their weights to compute the final score.

2.3.1 Gender weight

As the gender can be male or female or not specific so the weight of gender can be considered as

G_w	*Resume gender match JD gender*	$G_w = 1$
	Resume gender not match JD gender	$G_w = 0$
	JD gender not specific	$G_w = 1$

G_w *weight of gender*

2.3.2 Address weight

Practically, the job description tends to specify the preferred provinces because of the location of the company. This factor is also influenced to the candidate's decision. Due to the real-world practices, we only consider the province in the address.

A_w	*Resume province match JD province*	$A_w = 1$
	Resume province not match JD province	$A_w = 0$
	JD province not specific	$A_w = 1$

A_w *weight of address*

2.3.3 Degree weight

In this section, the weight of degree can be considered the highest degree in candidates' education.

D_w	*Resume degree match JD province*	$D_w = 1$
	Resume degree not match JD province	$D_w = 0$
	JD degree not specific	$D_w = 1$

D_w *weight of degree*

2.3.4 Languages weight

In this section, the weight of language can be considered as below

L_w	*Resume languages match JD gender*	$L_w = 1$
	Resume languages not match JD gender	$L_w = 0$
	JD languages not specific	$L_w = 1$

L_w *weight of languages*

2.3.5 Final scoring and ranking

After computing all the scores, we summarized the rescore similarity grading formula as formula (3) below. The final score is sorted in descending order, which means the resumes with the highest final scores are presented in descending order in the talent pool.

$$\text{Final score} = \sum \text{simi}(\vec{J}, \vec{R}), G_w, A_w, D_w, L_w \quad (3)$$

$\text{simi}(\vec{J}, \vec{R})$ *the similarity between vectors of job and resume*

3. Dataset and result discussion

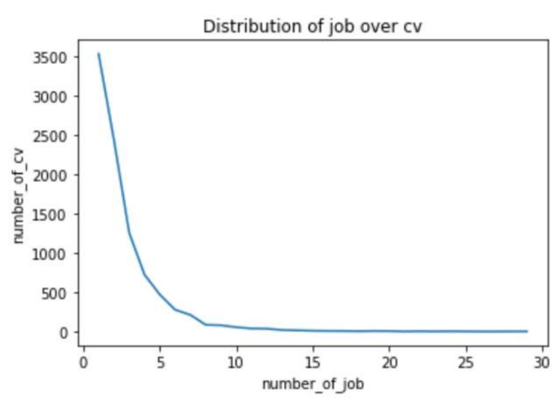


Figure. 3. Distribution of jobs in resume

The dataset for this proposed framework was collected from online recruiting channels in 2020, including 2,302 job descriptions in related to IT industry occupational categories., and 8,529 resumes. The set of resumes and the set of job descriptions are separated. The text language is limited to English. The distribution of jobs over resume is shown in Fig. 3. This statistic reveals the number of jobs in each resume. Most of the dataset consists of resumes containing from one to five job. For the rest, the resumes with number of jobs from six are not over 500 resumes.

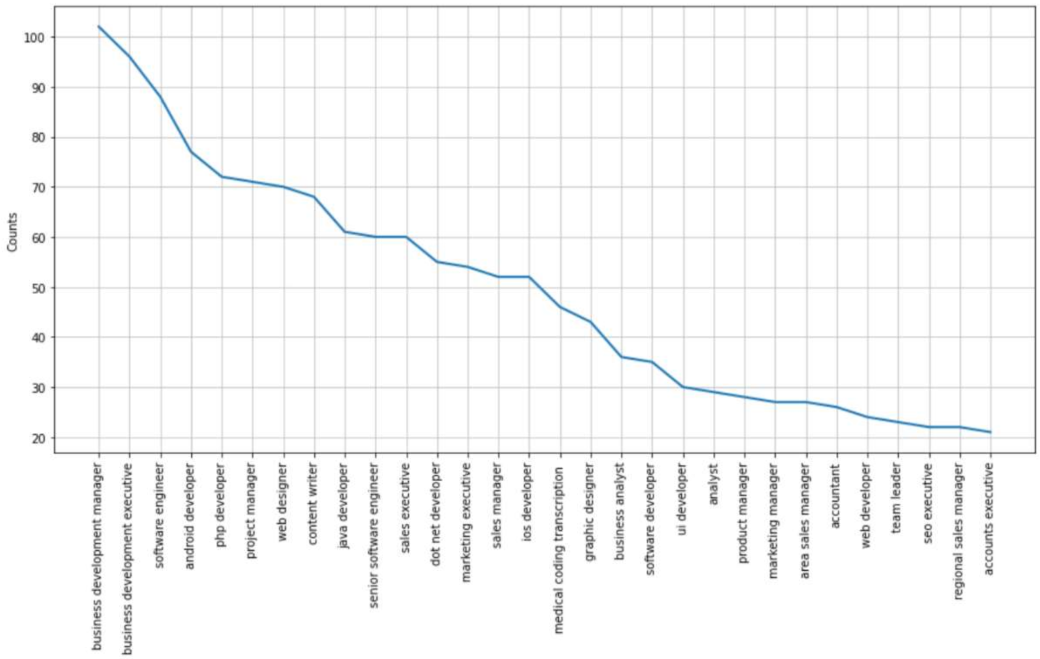


Figure 4. Common job titles in job post dataset

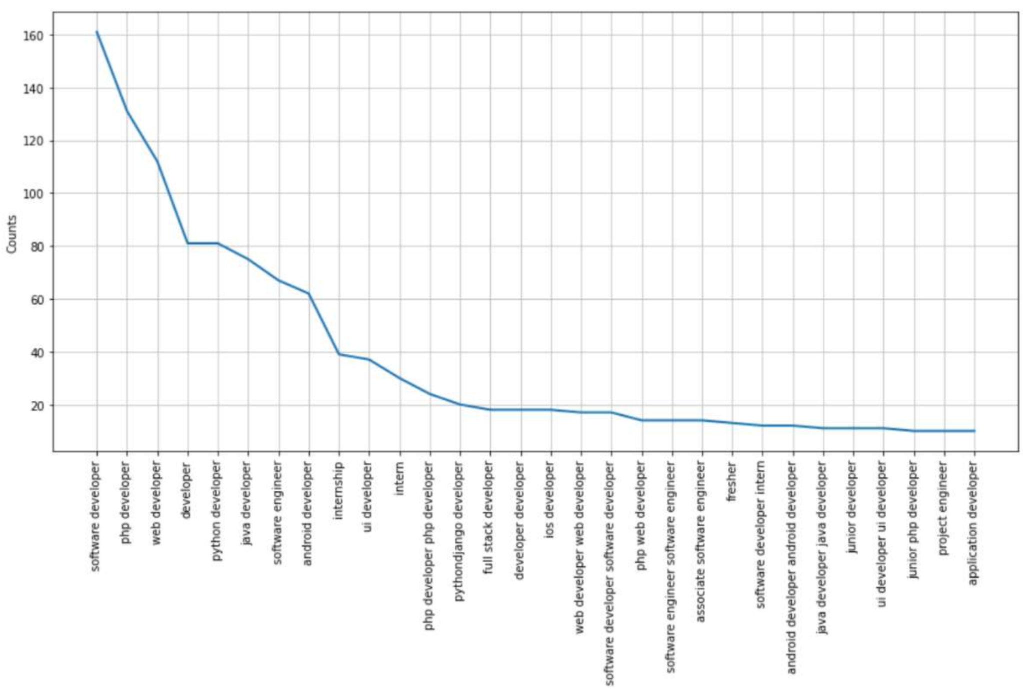


Figure 5. Common job titles in resume dataset

Furthermore, to avoid the mismatch in recommendation, the job titles in both resume dataset and job post dataset are taken into account. In this paper, for the sake of simplicity and public data available, only categories in IT industry are proceed. Shown in Fig. 5. and Fig. 6.

Fig. 6. shows the example result of the resume extraction part. In this information retrieval part, not only the personal information but the professional experience is also obtained. The results are stored in a separated dataset which might be worked on the future research.

```
NAME                          -
DESIGNATION                   - Senior Software Analyst
COMPANIES WORKED AT           - Accenture
LOCATION                      - Bengaluru
EMAIL ADDRESS                 -
DESIGNATION                   - Test Lead
DESIGNATION                   - Senior Software Analyst
COMPANIES WORKED AT           - Accenture
LOCATION                      - Bengaluru
COMPANIES WORKED AT           - Accenture
LOCATION                      - Bengaluru
DEGREE                        - B.TECH. in Information Techn
COLLEGE NAME                  - Amity University
SKILLS                        - QA (7 years), TESTING (6 year
```

Figure 6. Resume extraction

4. Discussion and Future Research

In this paper, a feasible framework for enhancing the talent pool was recommended. Firstly, the proposed framework pre-processes the text of resumes and job descriptions and performs personal information extraction. Secondly, the text-processed resumes and job descriptions are embedded and classified into two segments, then performed similarity measurement between resumes and job descriptions. Finally, to increase the talent pool efficiency, personal information extraction weights are added to calculate the final scores. The expected output is the list of resumes sorted grades in descending order in each occupational category.

The proposed framework is under implementation and employing but constructs a strong baseline to fill the pooling data's critical gap. As the talent pool has a strategic role in workforce planning, this framework expectedly aids the recruiters in reducing the workload and enabling them to access the efficient process of staffing.

There are some limitations in the paper. First, the future paper is still in experimental analysis based on an empirical case study to demonstrate the proposed framework's effectiveness. Second, future research should enhance the datasets and reduce manual tasks in preparing training to increase the framework's efficiency and accuracy.

Reference

Velay, M., & Daniel, F. (2018). Using NLP on news headlines to predict index trends. *arXiv* preprint arXiv:1806.09533.

Kong, S. H., Tan, L. M., Gan, K. H., & Samsudin, N. H. (2020, April). Fake News Detection using Deep Learning. *In 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics* (ISCAIE) (pp. 102-107). IEEE.

Bourgonje, P., Schneider, J. M., & Rehm, G. (2017, September). From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism* (pp. 84-89).

Yıldırım, S., Jothimani, D., Kavaklıoğlu, C., & Başar, A. (2018, December). Classification of" Hot News" for Financial Forecast Using NLP Techniques. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 4719-4722). IEEE.

Kaufman, D. R., Sheehan, B., Stetson, P., Bhatt, A. R., Field, A. I., Patel, C., & Maisel, J. M. (2016). Natural language processing–enabled and conventional data capture methods for input to electronic health records: a comparative usability study. *JMIR medical informatics*, **4**, e35.

Byrd, R. J., Steinhubl, S. R., Sun, J., Ebadollahi, S., & Stewart, W. F. (2014). Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. *International journal of medical informatics*, **83**, 983-992.

Jung, Y., & Suh, Y. (2019). Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. *Decision Support Systems*, **123**, 113074.

Menon, V. M., & Rahulnath, H. A. (2016, September). A novel approach to evaluate and rank candidates in a recruitment process by estimating emotional intelligence through social media data. In *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)* (pp. 1-6). IEEE.

Amin, S., Jayakar, N., Sunny, S., Babu, P., Kiruthika, M., & Gurjar, A. (2019, January). Web Application for Screening Resume. In *2019 International Conference on Nascent Technologies in Engineering (ICNTE)* (pp. 1-7). IEEE.

Gopalakrishna, S. T., & Vijayaraghavan, V. (2019). Automated Tool for Resume Classification Using Semantic Analysis. *International Journal of Artificial Intelligence and Applications (IJAIA),* **10**.

El Mohadab, M., Bouikhalene, B., & Safi, S. (2018). Automatic CV processing for scientific research using a data mining algorithm. *Journal of King*

Saud University-Computer and Information Sciences.

Zaroor, A., Maree, M., & Sabha, M. (2017, November). JRC: A Job Post and Resume Classification System for Online Recruitment. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 780-787). IEEE.

Bhatia, V., Rawat, P., Kumar, A., & Shah, R. R. (2019). End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT. *arXiv* preprint arXiv:1910.03089.

MULTI-RESOURCE ROUTING AND SCHEDULING PROBLEM-A CASE STUDY OF HOME HEALTHCARE SERVICE IN HO CHI MINH CITY

Trinh Minh Dang[1], Nguyen Hoang Huy[2]

[1,2]School of Industrial Engineering and Management, The International University - Vietnam National University, Vietnam, Email: [1]trinhminhdangknh@gmail.com, [2]nhhuy@hcmiu.edu.vn

Abstract: In-home healthcare services, patients with different health condition backgrounds require different types of resources, and the serving time among these patients also varies. The hospitals and their vehicle fleet with limited carrying capacity are responsible for transporting medical equipment, medicine, medical staff to patient locations. The problem is categorized as Multi-resources Routing and Scheduling problem (MRRSP). A Mixed-Integer Linear Programming model is developed to decide the scheduling and routing decision of the vehicle carrying the resources from the hospital to patient locations that minimize the total weighted completion time by choosing the shortest distance among the visiting locations as well as considering the processing time at these points. The MRRSP is an NP-hard problem and cannot be solved within a polynomial-time, so a heuristics algorithm is developed to solve the real-size problem. The computational time of the Greedy algorithm proves to be effective when applying in real-life situations.

Keywords: Multi-resources Routing and Scheduling problem, Mixed-Integer Linear Programming, Greedy Algorithm, Logistics, Home healthcare.

1. Introduction

1.1. Home healthcare service

In Vietnam, especially in major cities such as Ho Chi Minh City or Hanoi, the demand for high-quality healthcare services is increasing rapidly. Meaning that more and more patients are willing to pay more to receive a better healthcare service. During the last decade, a lot of doctors, physicians from governmental hospitals have been running private clinics after working hours, and this option of receiving medical treatment has become popular among patients ever since. The advantages of private clinics can be listed as short waiting time of customers for their serving and customers can find and visit those private clinics run by well-known, experienced doctors, who work for state-own hospitals with high reputation. However, not all of these doctors open their clinics, or because of their reputation and popularity among patients, their clinics may also experience an overload similar to the hospitals because these private clinics operate during a particular period of the day.

According to a survey conducted in 2017 by Seoul National University researchers (Bang et al., 2017) about the need for home healthcare service of the elder's age from 60 or above in Quoc Oai district-Hanoi. The result indicates that more than 21% responded that they need home healthcare service which a wide range of purposes such as managing diseases, health-promoting, rehabilitating. Because Quoc Oai is a promote district from Hanoi city centers and the limitation in targeting the group of elder patients only, the demand for home healthcare service is expected to increase significantly for other groups of patients within other geographic locations as well. Therefore, it is important for hospitals, especially private hospitals to adapt to the trend and provide proper caring services to the customers. Private hospitals can benefit based on the generated results of this study to improve their service serving systems which regards the provision of home healthcare service to the customers.

Bringing health care to people in their homes can have significant benefits—reduces costs, adds convenience (particularly for the elderly patients who have difficulty visiting the doctor's office). The case study of the Washington Hospital Center in Washington, DC, conducted by the National Research Council (U.S) Committee on the Role of Human Factors in Home Health Care (2011) discusses the popularity of home healthcare service in the US over the conventional healthcare service at other medical facilities. The hospital has been providing home healthcare service for more than a decade, serving nearly 600 patients. Their program indicates that a significant amount of cost has been reduced for the patients who participate in the program. Moreover, health care is increasingly occurring in residential contexts rather than in professional medical environments such as hospitals, medical clinics, rehabilitation camps.

1. The costs of providing health care at formal medical facilities are increasing.

2. The prevalence of chronic conditions across the entire age spectrum is growing (particularly conditions related to obesity, such as diabetes), and growing along with it is the demand for health care.
3. Some types of health care professionals are in short supply, which shifts the burden of some types of care onto lay caregivers to fill the gap. (In hospitals, professional doctors, nurses, medical staffs will process patients from many kinds of illnesses, a large number of these illnesses can be transferred to caregivers to relieve the workload for those staffs at the hospitals)
4. Consumers want to be independent in their health management and are seeking more home-based services.

The *tasks* involved in home-based health care include activities for maintaining health, activities associated with episodic care (e.g., in response to illness or injury) or chronic care, or activities to support the end of life. The medical conditions involved may be simple and involve little time and no medical equipment, or they may be very complex and consume many hours every day and require the use of complicated devices.

The *devices and technologies* for health care in the home cover a vast range, from simple first aid tools to respiratory equipment, and from meters and monitors to computer equipment and software associated with interconnected electronic systems. Some of this equipment was designed only for professional use but is finding its way into the home, nevertheless.

These activities include disease prevention and health maintenance activities (e.g., regular exercise, health information seeking) and management of acute ailments (e.g., colds, minor infections, injuries) and chronic diseases (e.g., diabetes, multiple sclerosis) and disabilities (e.g., vision, hearing, mobility, or cognitive impairment). One large population of people who receive care at home are those with chronic conditions, such as hypertension, asthma, diabetes, cancer, HIV/AIDS, chronic respiratory disease, neuromuscular disease, dementia, or emotional disorders.

The hospital's target customers are mainly the elder and patients who have difficulty in visiting the hospital frequently and need medical aid in their living space. There are mainly 3 types of patients that the hospital receives:

- Periodic visit: Patients that require the home healthcare service weekly, monthly.
- Pre-booked visit by a doctor: An appointment is made by the doctor for a particular patient when patients' blood, urine, is taken at the first visit for testing purposes and the results are returned to patients in the next visit.
- Pre-booked visit by a patient: Appointment made by the patient for doctor visit on a specific day.

For periodic patients, the visiting schedule is fixed and unchangeable, meaning that there is a visiting time agreement made between the periodic patients and the hospital. Moreover, periodic patients may request a specific doctor and/or nurse that can keep track of their health weekly, monthly, and so on. For this type of patient, the mid-term to long-term visiting plan can be made based on the already established schedule and the availability of the appointed doctors/nurses. For that reason, the assigning problem is narrowed down to specific human resources in a specific time window, and the schedule can be planned easily and quickly a week or a month before the visiting day because the visiting activity is repetitive.

This paper aims to solve an existing problem in the home healthcare industry. Besides periodic patients, some pre-booked patients request the service on a specific day. Besides, some patients have appointments with doctors for a re-visit after doctors receive their testing results. The booking period may occur at a half or most a day before the visiting period, the patient's lists may be adjusted at the beginning of every working shift to adapt flexibly with the current situation. Therefore, daily planning or short-term planning is made by the hospital due to the consideration of the heterogeneous constraints or the consideration of the unexpected patient's admission. Moreover, the working schedule for the medical staff is also different, therefore the plan is made for every working shift and is based on the current availability of staff on that particular shift.

For periodic patients, the expected completion time may be true for most cases as a patient's health condition is well known by doctors/nurses and hospital staff can perceive how long will the service take place. In contrast, for a new case such as pre-booked patients in which patient information is provided partly or inadequate, unexpected extra time may arise in case of new health-related problems are detected or samples for testing purposes such as (blood, urine) are required to be taken. To reduce the arising of unwanted time which may affect the whole visiting plan, patients are categorized by priority factors. The higher the priority is, the more likely it is that patient is scheduled to be visited as soon as possible.

1.2. Problem statement

In the scheduling phase, the amount of both resource types supplied to demand points will be determined and assigned to the vehicle fleets stationing at hospitals. The decisions for which resources to be sent to which locations on which vehicles are taken into consideration to minimize the total completion time of the medical team. Since more than 1 type of resource is required to be delivered to multiple sites, the supply chain network becomes more complicated when it comes to scheduling multiple resources Lei et al. (2015). Adding to the complexity of the problem, hospitals are subject to resources limitation constraints such as distribution centers capacity, vehicle capacity, a finite number of the vehicle in each fleet, number of medical staffs on the shift. After finishing the scheduling phase, the vehicles must transport resources to the assigned demand points as soon as possible. Therefore, the routes for the vehicle must be determined in a way that minimizes the total traveling time. The visiting sequence which is determined not only by considering the traveling times but also the weighted factor will be considered to find the optimal route. Demand locations are usually geographically dispersed and vary from point to points in terms of service processing time. The waiting time may impact significantly on the total traveling time of the vehicle and the longer the vehicle must wait to depart.

1.3. Objectives

The Home Healthcare Service routing and scheduling problem considered in this paper is categorized as Multi-resources Routing and Scheduling problem (MRRSP). The visiting plan for the medical staff needs to be made with the decisions on the travel routes and arrival times of the medical staff while minimizes the total completion time of the whole operation and meets all of the patient's demands while utilizing the limited resources. This research aims at developing a mathematical model and a framework to solve the MRRSP with an outcome of minimizing the total completion time with the decisions in scheduling both types of resource, assigning and routing vehicles for all the demand points within an acceptable computing time.

2. Literature review

2.1. Vehicle Routing Problem in healthcare sector

Multi-resource routing and scheduling problems can be categorized as logistics and transportation problems. Therefore, the related works were considered which take into many elements such as routing and scheduling. Before this study, there is a substantial amount of work considering the vehicle routing and/or scheduling for logistics using a mathematical programming approach. Among those research, some deal with scheduling problems, some deal with routing problems, and many papers consider both problems in one context. Pillac, Van Hentenryck, and Even (2016) considers the disaster scenario of evacuating which the main objective is to minimize the evacuee flows whereas the sub-problem is to generate more evacuation plans by finding new evacuation paths. A similar disaster scenario is also studied by Bodaghi, Palaneeswaran, et al. (2020) in the case of bushfire evacuation by providing routing solutions for vehicles to evacuate victims out of the affected areas with time windows and uncertain risks. Bruni, Beraldi, and Khodaparasti (2018) work resemble this study by considering the vehicle scheduling and routing which carry supplies for the victims. In this study, the authors emphasize the importance of relief vehicle's arrival time over the vehicles routing cost minimization, which is the main focus of many studies on the same field.

In the logistics field which regards healthcare services, Shi, Boudouh, and Grunder (2019) study the home healthcare routing and scheduling, and propose a non-deterministic model to cope with the uncertainties in terms of vehicle traveling times and customer's processing time. This study deals with the problem of the variety in customer demand who require many forms of medical treatments served by the caregivers. Considering the limited resource of caregivers, the model addresses the problem of skill-labor assignment and vehicle routing. Zhu et al. (2019) address the victim's routing problem which considers the severity factor. In this study, 2 parallel models are promoted, each of them covers a realistic scenario which are identical injury or diverse injury degrees. The objective is to find a possible vehicle routing solution that may optimize the operational cost while takes into consideration the humanitarian factors on which victims should be treated equally or which ones should be prioritized over the other ones.

2.2. Open Vehicle Routing Problem (OVRP)

Among many sub-problems of VRP, many researchers focus on the Open VRP (OVRP). The main difference is that vehicles are not required to return to the depots that they station in, vehicles are allowed to move to the next locations and may return to the depot once they reach the final destinations, although this final route is not critical. Bruni, Beraldi, and Khodaparasti (2018) consider the same OVRP problem of routing the vehicles to transport critical supplies to the disaster victims at the affected areas where they deicide the vehicle acquisition decisions, scheduling decisions, and routing decisions to minimize the total transportation time, number of required vehicles and so on.

2.3. Multi-resource Routing and Scheduling problem (MRRSP)

The above research successfully addressed and solved the routing scheduling problems, however, among those research, very few considered the optimization model which is subject to multiple resources. In most cases, expendable resources and non-expendable resources are studied individually and only a few exceptions have works done on both types of resources. Lee et al. (2013) proposed a resource scheduling model that consists of expendable resources and non-expendable resources and the availability of both resources to deliver the service at customer's sites is strictly required. Another resemble research is Lei et al. (2015), the model involves the traveling of the medical teams and the distribution of the medical supplies. This relates to the problem of dispatching time of the expendable resources from the distribution centers by the medical team, denoted as non-expendable, to minimize the total tardiness of the whole operation. Most studies develop a mixed-integer programming model and applied heuristics algorithms to assign and schedule the allocation of both resources in the supply chain network. In most cases, the routes are pre-determined, and this assumption may be invalid in realistic cases, the routing problem should be addressed, and it falls into the vehicle routing problem (VRP). However, as mentioned above, there is a lack of research that considers the routing and scheduling problem for multiple resources. Bodaghi, Shahparvari, et al. 2020) study is among a few ones that consider the addressed research gap, they construct a mathematical model inspired by the case of allocating and assigning expendable resources and non-expendable resources to designated vehicles. The routing decision is also considered to minimize the total weighted completion time.

2.4. Mix-integer programming model for MRRSP

The mathematical model for MRRSP is usually a mixed-integer programming model where decisions are made following a set of pre-determined constraints. The exact solution such as the Branch-and-cut algorithm is applied to solve the mixed-integer programming model. Moreno, Munari, and Alem (2019) use this method to solve the Crew Routing and Scheduling Problem to minimize the inaccessible time of the demand points while waiting for the crew to fix the roads connecting the depot with these demand points. The objective can be achieved based on the proper decisions on (i) the scheduling of the repairing crew, (ii) the paths between 2 successive damaged nodes, (iii) the paths between the depot and the demand nodes. Bertazzi et al. (2019) develop a three-phase metaheuristic to solve the Multi-Depot Inventory Routing Problem (MDIRP). The problem is formulated as a mixed-integer linear programming model and used the Branch-And-Cut algorithm to obtain the solution. The solution is then compared with the one generated from the metaheuristic. Ho et al. (2008) approach the multi-depot vehicle routing problem (MDVRP) using a meta-heuristic Genetic Algorithm (GA) which is widely used to solve transportation problems that relate to VRP. The objective is to draw an effective routing and scheduling scheme to distribute finished goods from depots to customers within the shortest amount of time. Bodaghi, Shahparvari, et al. (2020) formulate a Multi-Resource Scheduling and Routing problem (MRRSP) for emergency relief which regards the problem of delivering multiple types of resources to customers in the shortest time. Along with the solution from the exact method, this paper also develops a Greedy Algorithm to obtain a near-optimal solution within an acceptable computational time for medium to large size problems. Wex et al. (2014) study also focuses on the emergency logistics operation of scheduling and routing vehicles to transport necessity to affected areas, customers demand points to minimize the total completion time which takes into consideration the weighted destruction factor. A Greedy Algorithm is developed to solve the problem as it is proved to be NP-hard and cannot be solved in an acceptable computing time, noting that under the emergency logistics situation, time responsiveness plays a crucial role, and decisions are expected to be made as soon as possible, so there comes the need for a method that can generate solution quickly, yet accurately.

3. Methodology

3.1. Mathematical model

The mathematical formulation that will be used in this paper will be based mainly on the study of Bodaghi, Shahparvari, et al. (2020) on solving the Multi-resource Routing and Scheduling problem (MRRSP).

Notation

V: *Set of vehicles.*
V_k: *Set of vehicles of hospital k.*
I: *Set of n demand points {0, n+1} whereas 0 and n+1 represent the starting and ending nodes.*
K: *Set of hospitals.*
G: *Set of all resource types.*
T: *Set of expendable resources.*
R: *Set of non-expendable resources.*
i, j, l: *Indices for demand points,* $i, j, l \in \{0,1, \ldots, n, n+1\}$
k: *Index for hospital,* $k \in K$
t: *Index for resource type*

Parameters

d_i^t: *Quantity of a specific resource type t required by a demand point* i
cap_t^v: *Capacity of vehicle v for carrying resource* t
cap_k^t: *Capacity of the hospital for allocating resource t*
$travel\_time_{ij}$: *Time required to travel from location* i *to* j
$travel\_time_{ki}$: *Time required to travel from hospital* k *to location* i
p_i^t: *Processing time of non-expendable resource* t *at location* i, $t \in R$
w_i: *Weighted factor at location* i
M: An arbitrary big positive number

Decision variables

$$X_{ij}^v \in \{0,1\}$$

$$\begin{cases} 1: if\ vehicle\ v\ from\ i\ transports\ resources\ to\ j \\ 0: otherwise \end{cases}$$

$$Y_{ki}^v \in \{0,1\}$$

$$\begin{cases} 1: if\ hospital\ k\ supplies\ i\ with\ vehicle\ v \\ 0: otherwise \end{cases}$$

$quantity_{kv}^t$: *Total quantity of resource type* t *from hospital* k *transported by vehicle* v
S_i: *Starting time of the treatment operation at location i*
C_i: *Completion time of the service operation at location i*
P_i: *Total processing time of the non-expendable resources at location* i

Objective function

$$Minimize\ Z = \sum_{i=1}^{n} w_i\, C_i$$

Constraints

$$\sum_{j=1}^{n+1} \sum_{v=1}^{V} X_{ij}^v = 1 \quad \forall i \in I/\{n+1\}\ \&\ i \neq j \qquad (1)$$

$$\sum_{j=1}^{n+1} X_{0j}^v = 1 \qquad \forall v \in V \qquad (2)$$

$$\sum_{v=1}^{V} X_{ij}^v + \sum_{v=1}^{V} X_{ji}^v \leq 1 \quad \forall i, j \in I\ \&\ i \neq j \qquad (3)$$

$$\sum_{i=0}^{n} X_{il}^v - \sum_{j=1}^{n+1} X_{lj}^v = 0 \quad \forall v \in V, \forall l \in I/\{0, n+1\}\ \&\ l \neq i\ \&\ l \neq j \qquad (4)$$

$$\sum_{k=1}^{K} \sum_{v=1}^{V} Y_{ki}^v = 1 \qquad \forall i \in I/\{0, n+1) \qquad (5)$$

$$quantity_{kv}^t = \sum_{i=1}^{n} Y_{ki}^v\, d_i^t \quad \forall k \in K, \forall v \in V, \forall t \in T \qquad (6)$$

$$quantity_{kv}^t = \max_{(i \in I/\{0,n+1\}}\, (Y_{ki}^v d_i^t) \quad \forall k \in K, \forall v \in V, \forall t \in R \qquad (7)$$

$$quantity_{kv}^t \leq cap_t^v \ \ \forall k \in K, \forall v \in V, \forall t \in G \qquad (8)$$

$$quantity_{kv}^t \leq cap_k^t \ \ \forall k \in K, \forall v \in V, \forall t \in G \qquad (9)$$

$$P_i = \sum_{t=1}^{R} p_i^t \qquad \forall i \in I/\{0, n+1) \qquad (10)$$

$$travel\_time_{ki} \times Y_{ki}^v \leq S_i + M(1 - X_{0i}^v) \quad \forall i \in I/\{0, n+1), \forall v \in V \qquad (11)$$

$$S_i + P_i + travel\_time_{ij} \leq S_j + M(1 - X_{ij}^v) \quad \forall v \in V, \forall t \in R, \forall i, j \in I\ \&\ i \neq 0\ \&\ j \neq (n+1)\ \&\ i \neq j \qquad (12)$$

$$C_i \geq S_i + P_i \qquad \forall t \in R, \forall i \in I \qquad (13)$$

$$Y_{ki}^v = 0 \ \ \forall i \in I,\ \forall k \in K,\ \forall v \notin V_k \qquad (14)$$

$$C_i \geq 0;\ S_i \geq 0;\ P_i \geq 0;\ quantity_{kv}^t \geq 0 \quad \forall v \in V, \forall k \in K, \forall i \in I, \forall t \in G \qquad (15)$$

The objective is to minimize the total weighted completion time for all demand points (except the dummy nodes). Constraints (1) to (4) relate to determining the sequence of the demand points visited by each vehicle to deliver the required resources. Two additional dummy points (the starting and ending points) are defined for each vehicle (denoted by "0" for the starting point and "n+1" for the ending point).

Constraint (1) guarantees that there is exactly one demand point j instantly visited after the demand point i by any vehicle to deliver resources. Constraint (2) ensures that each vehicle initially starts from its starting point (i = 0). Constraint (3) eliminates the loops involving two nodes in processing the demand points by a particular vehicle, ensures that the vehicle can move in one direction only. Constraint (4) creates the visiting sequence for a vehicle, guarantees that if a previous node l is available for each demand point (demand point j), and each vehicle, it would be a successive node for a previous demand point (demand point i). The sequence ends at the dummy node 0.

Constraint (5) establishes the relationship between a demand point, the vehicle, and the hospital from where the vehicle carried the resources, ensures that each demand point receives resources from hospital k, carried vehicle v.

Constraint (6) calculates the total quantity of expendable resources delivered from the hospital by vehicle. Constraint (7) calculates the quantity of non-expendable resources by vehicle.

Constraint (8) ensures that the quantity of resource type delivered by a vehicle does not exceed its capacity. Similarly, constraint (9) ensures that the quantity of resource type delivered by a vehicle of the hospital does not exceed that hospital capacity for that resource type.

Constraint (10) calculates the total processing time of all non-expendable resources when delivering service at a demand point.

Constraint (11) to (12) are time-related constraints. Constraint (11) ensures that the starting time of the visit at each demand point is not earlier than the arrival time of resources from the hospital with the vehicle if the vehicle's current is at the hospital. Constraint (12) ensures that the starting time of the visit is not earlier than the arrival time of the resources from the previous demand points if the vehicle's current location is at other nodes rather than the hospital. Constraint (13) calculates the completion time of the healthcare service at the demand point.

Constraint (14) guarantees that the only vehicle from the hospital can deliver resources of that hospital to the demand points, otherwise, the decision variable will always equal 0. Constraint (15) sets the domains for the defined variables.

3.2. Case study

Due to the privacy policy of the hospital, "Hospital K" will be addressed instead of using the real hospital name. Hospital K is established in 2009 with its mission of providing high-quality services to patients such as health examination on patient's request, enterprise packages, testing packages, home healthcare service. For home healthcare service, the hospital brings to patients many necessary services ranging from periodic health examination, chronic diseases (diabetes type 1,2, high blood pressure, asthma) caring, physical therapy. A team of medical staff is formed consists of 1 doctor and 1 nurse; each team will visit the assigned patients once. Doctor and nurse may perform medical services individually and the service time per patient is projected from 20 to 40 minutes per nurse/doctor. The dataset is composed of hospital K and its sub-branch, which is also under the management of hospital K, each hospital owns a fleet of 7 vehicles. There are 42 patients who request the home healthcare service on a specific morning shift. The patients are characterized by different features: addresses, health conditions, demands different types of resources, priority level. The visiting plan consisting of visiting sequence and the arrival time of each medical times is established for the hospital to make appointments with the patients, assumed that all patients are agreed to receive the healthcare service on the morning shift starting from 7:00 AM ($t_0 = 7$). The operation of the healthcare service is illustrated in detail in Figure 1.

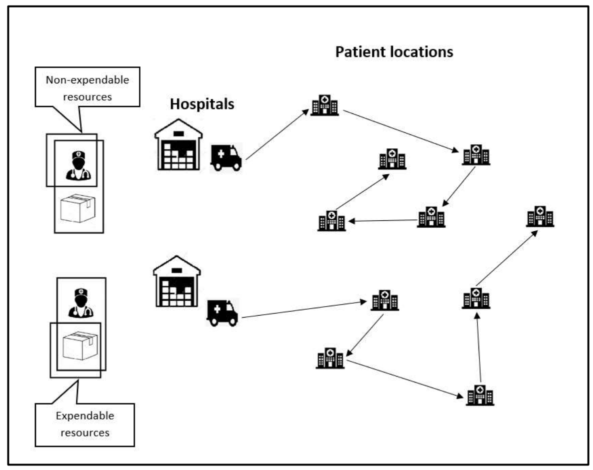


Figure 1: Home healthcare service operation

3.3. Greedy Algorithm

3.3.1 Scheduling Phase

The decision-making process is made based on considering the patient's demands for different types of resources and the availability of these resources at the hospital as well as the carrying capacity of the vehicle for these resources from its hospital to the patient's location. The procedure is described below step by step and by the flow chart in figure 2.

- **Step 01:** Consider a demand point (patient location)
- **Step 02:** Check the availability of all the resources at the hospital, if one out of any resource quantity is not available for allocating, move to the next hospital.
- **Step 03:** If the considered hospital is capable, check the capacity of vehicles at that hospital for carrying those required resources. If the whole fleet of one hospital are all infeasible for transporting resources to that patient, move to the next hospital and repeat **Step 02**

- **Step 04:** If both hospital and the vehicle are capable of supplying and transporting resources, the allocating decision is made
- **Step 05**: A list of capable nodes to that patient location is established. The capable nodes can be:
  - Hospital: if the current location of the vehicle is at the hospital
  - Previous patient location: if the current location of the vehicle is at another patient location other than the considered one.

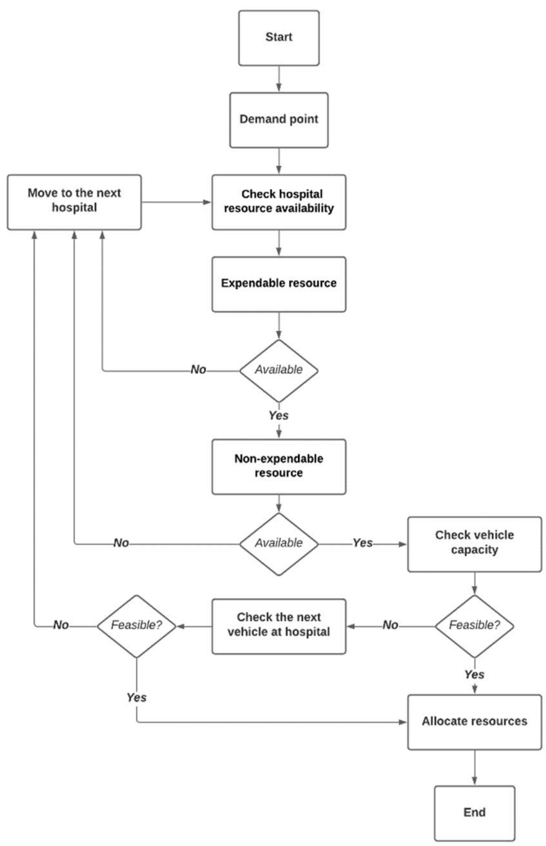


Figure 2: Illustration of the resources scheduling phase.

3.3.2. Routing Phase

The VRP model needs to serve a set of customers, either concentrated in the same place or scattered in different places, with a fleet of vehicles. The items for each customer must be delivered in one visit by one vehicle. The second phase aims to generate a restricted number of possible routes for the vehicles and choose the one with the lowest time value. Based on the capable nodes list obtained in the Scheduling phase, there are a set of vehicles that are capable of transporting resources to the considered location as well as these vehicles' current locations. For the total time value, 2 elements need to be considered which are:

- Traveling time: the time required to travel from the previous capable nodes to the considered node, denoted as $\boldsymbol{travel\_time_{(N_i)i}}$ where i is the considered location.
- The completion time of the vehicle: as waiting time for the vehicle to complete its previous visit may affect the total completion time even if the traveling time from the current location of that vehicle to the consider node is the lowest. The completion time of the vehicle is denoted as $\boldsymbol{C_v}$

The total time value equals the sum of traveling time and the completion time of the vehicle:

$$\textbf{Total time} = \boldsymbol{travel\_time_{(N_i)i} + C_v}$$

The figures below illustrate how the time value is calculated based on 2 scenarios which are:

- The vehicle's current position is at the hospital.
- The vehicle's current position is at another location rather than the considered location.

In figure 3, none of the patients is assigned to any vehicles so the current position of all vehicles is at hospital k. There is only 1 capable node to patient 1 which is the hospital. The traveling time of all vehicles from this capable node to patient 1 is equal (travel distance between the hospital and patient 1) and the completion time of all vehicles is 0. Therefore, the total time of all the elements in the capable nodes list now is equal to each other. Scenario 1 is described shortly as follow:

$$\boldsymbol{C_v = 0}$$

$$\boldsymbol{travel\_time_{(N_1)1} = travel\_time_{k-1}}$$

$$\textbf{Total time} = \boldsymbol{travel\_time_{(N_1)1} = travel\_time_{k-1}} \quad \text{(equals for all vehicles)}$$

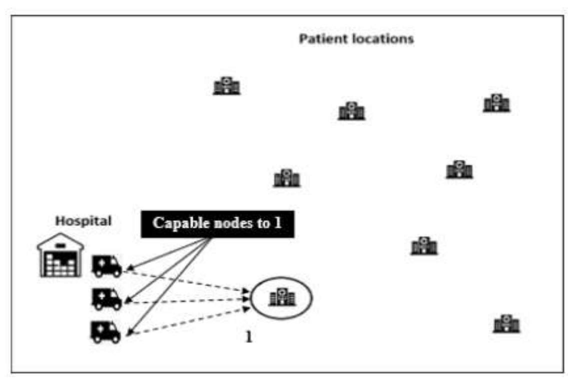


Figure 3: Illustration of the routing phase case 1

In figure 4, the capable nodes to patient 1 are the hospital, patient 3 location, patient 2 location, respectively. The completion time of the vehicle at the hospital is 0, however, the completion time of the vehicle at locations 3 and 2 are now updated by traveling time from the hospital plus processing time $(\boldsymbol{p_n})$ at the current location if the previous nodes are at the hospital. If the vehicles traveling to location 2 or 3 initially depart from a location rather than the hospital, the completion time is now calculated as the total amount of the completion time and the processing time of the previous location and traveling time from the previous location. Therefore, the total time is now different from the capable nodes when adding the calculated completion time of every single point in the capable nodes with the traveling time from them to the considered patient 1. Scenario 2 is described shortly as follow:

- Vehicle a's current location is at the hospital.
  $\boldsymbol{C_a = 0}$
  $\boldsymbol{travel\_time_{a-1} = travel\_time_{k-1}}$

- Vehicle b's current location is at patient 2.
  $\boldsymbol{C_b = travel\_time_{k-2} + p_2}$
  $\boldsymbol{travel\_time_{b-1} = travel\_time_{2-1}}$

- Vehicle c's current location is at patient 3.
  $\boldsymbol{C_c = travel\_time_{k-3} + p_3}$
  $\boldsymbol{travel\_time_{c-1} = travel\_time_{3-1}}$

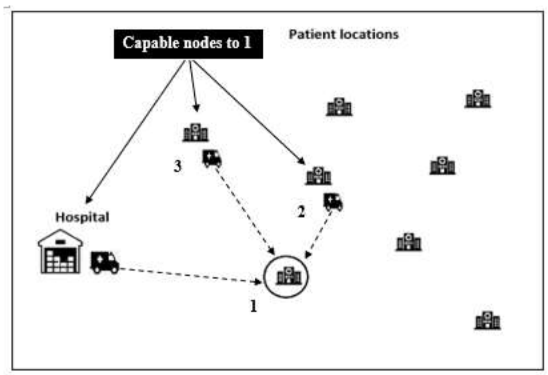


Figure 4: Illustration of the routing phase case 2

3.3.3. Solving procedure

Greedy algorithm pseudo code for solving the similar problem developed by Bodaghi, Shahparvari, et al. 2020 is referred. The pseudo code for the greedy heuristics algorithm is described as below:

Phase 0: Preparation
Parameters:
$Set\ of\ patients\ i \in I$
$w_i: weighted\ values$
cap_t^v**:** capacity of vehicle v for carrying resource type t
cap_k^t**:** capacity of hospital k of supplying resource type t

Phase 1: Scheduling
Sort patient locations by decreasing w_i values.
$C_v \leftarrow \emptyset$: **Initiate** the completion time of the vehicle.
$starting\_time_i$
$\leftarrow \emptyset$: $\boldsymbol{Initiate}\ the\ starting\ time\ at\ patient\ i$
$C_i \leftarrow \emptyset$: **Initiate** the completion time at patient i
$S_v \leftarrow \emptyset$: **Initiate** the visiting sequence of the vehicle.
$vehicle\_cap_t^v \leftarrow cap_t^v$: **Initiate** the capacity of each vehicle.
$hospital\_cap_k^t \leftarrow cap_k^t$: **Initiate** the capacity of each hospital.
$N_i \leftarrow \emptyset$: **Create** a list of vehicles that are able to visit node $\boldsymbol{i}$ (capable nodes list)
for $\boldsymbol{i}$ = 1 to $\boldsymbol{n}$ do:
  if $curr\_hospital\_cap_k^t \geq d_i^t$
    for v in V_k:
      if $curr\_vehicle\_cap_t^v \geq d_i^t$:
 $N_i \leftarrow v$: Add the selected vehicle into the capable nodes list.
  if $N_i = \emptyset$:
  abort #Infeasible#

Phase 2: Routing
  $T_{N_i} \leftarrow travel\_time_{(N_i)\,i} + C_v$ (1)
  choose v: $v \leftarrow \{v \mid min(T_{N_i})\}$ (2)
  update: $C_v \leftarrow T_{N_i} + p_i^t$ (3)
  update: $S_v \leftarrow S_v \cup \{ i \}$ (4)
  update: $vehicle\_cap_t^v \leftarrow \{vehicle\_cap_t^v - d_i^t\}$
  update: $hospital\_cap_k^t \leftarrow \{ hospital\_cap_k^t - d_i^t\}$
  $Z_i = w_i * C_v$: *Calculate the weighted completion time at patient* i
$Z \leftarrow \sum_{i=1}^{n} Z_i$: Calculate the total weighted completion time.
Return $Z,\ S_v, C_v$

----

(1) *Calculate the total time of each vehicle in the capable nodes list to the considered node i*
(2) *Select vehicle* v *with the lowest total time value.*
(3) *Update the completion time of that vehicle v*
Add patient i to the visiting sequence of vehicle v.

3.4. Data collection

The data collected in this paper comes from a private hospital in Ho Chi Minh city in which they are providing home healthcare service to patients alongside the *conventional* healthcare service. The hospital has many facilities within Ho Chi Minh city and their patients also reside within the city area. The demand for each type of resource is obtained based on the patient's health conditions and these resources will be sent along with the team of doctors and/or nurses when they visit the patient. There are 3 types of input data, which includes:

- Customers demand per resource type.
- The capacity of hospitals per resource types.
- The capacity of vehicles per resource types.
- Processing time per patient.

To find the shortest route, information about the travel time needs to be obtained to determine the shortest path for every vehicle so that every location is guaranteed to be visited within the shortest amount of time. The list of patients is given by the home healthcare department of the hospital. Basic information of patients is provided such as gender, age, address, however, patient names will not be shown due to the confidential policy of the hospital, instead, they will be denoted by numbers. The addresses of customers are provided, and Google Maps is used as a tool to determine traveling distances between patients as well as between hospitals and patients.

4. Results

4.1.2. Computing time

The Heuristics described above was coded in Python, and the MILP model was coded in CPLEX studios. Computational experiments were carried out on a PC equipped with an Intel Core I5-8250 CPU running at 2.80 GHz, with 8 GB of RAM with Windows 10. This paper used the MIP solver IBM CPLEX 12.8.0 using its default settings. There are 5 datasets with the number of hospitals, medical team, and vehicle increases per dataset. The objective values and the computing time for all 5 datasets are described as followed:

Table 1: Different data sets results by exact method

	Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**	**Dataset 5**
Objective value	2356	2694	5439	10967	13021
CPU time(s)	5.52	17.39	287.63	14400	14400
Gap (%)	0	0	0	15.5%	21%

Table 2: Different data sets results by heuristics

	Datase t 1	**Datase t 2**	**Datase t 3**	**Datase t 4**	**Datase t 5**
Objective value	2505	2761	6593	12736	14563
CPU time(s)	0.12	0.22	0.43	0.78	1.23

The Greedy heuristic is a better approach when dealing with complex problems. The greedy heuristics can give a good solution within a reasonable time while the best solution from the exact is time-consuming to provide for cases when the exact method fails to provide the results within a polynomial time. Moreover, the objective value gaps of the two approaches are acceptable and heuristics can be used as an alternative problem-solving tool for the exact method. The result comparison of the two methods are illustrated in figure 5.

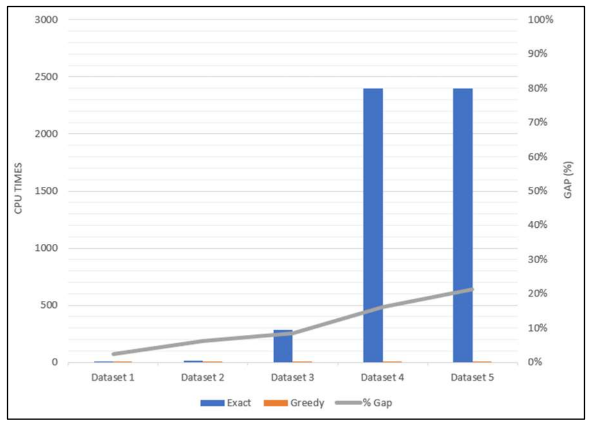


Figure 5: Result comparison

4.2. Greedy heuristic solution

The visiting sequences and the completion time of each vehicle are calculated and shown in table 3. In table 4, the generated starting time provides the arrival time of the medical staff at a particular location as well as the completion time at that site. Based on the arrival time, the appointment can be made in advance with the patients and also provide the medical staffs who take care of these patients, a time window for each patient to follow and accomplish their duty within the pre-determined time. The example illustration of the route of each vehicle with the starting and completion time is shown in figure 6 and 7.

Table 3: Visiting sequence and the completion time for the vehicle fleet.

Hospital	**Vehicle**	**Visiting Sequence**	**Completion time (hours)**
	Vehicle 1	20→21→26	3.1
	Vehicle 2	39→35→13	3.23
	Vehicle 3	23→40→8	2.87

1	Vehicle 4	31 →41→ 18	2.52
	Vehicle 5	19→ 24 →2	3
	Vehicle 6	36→ 22	2.4
	Vehicle 7	29→7→11	2.67
2	Vehicle 8	37→38→34	3.47
	Vehicle 9	4→28→16	2.63
	Vehicle10	5→9→27→10	3.17
	Vehicle 11	33→30→17	3.2
	Vehicle 12	12→ 6 →32	3.33
	Vehicle 13	42→14→ 15	2.95
	Vehicle 14	25 →3→1	3.07

Table 4: Appointment time and completion time at each patient location

Patient	Appointment time	Expected Completion time
1	9:30 AM	10:04 AM
2	9:15 AM	10:00 AM
3	8:34 AM	9:22 AM
4	7:13 AM	7:55 AM
5	7:13 AM	7:48 AM
6	8:22 AM	9:09 AM
7	7:57 AM	8:43 AM
8	9:07 AM	9:52 AM
9	7:49 AM	8:36 AM
10	9:18 AM	10:10 AM
11	8:55 AM	9:40 AM
12	7:33 AM	8:15 AM
13	9:21 AM	10:13 AM
14	8:42 AM	9:13 AM
15	9:16 AM	9:57 AM
16	8:48 AM	9:37 AM
17	9:19 AM	10:12 AM
18	8:46 AM	9:31 AM
19	7:28 AM	8:19 AM
20	7:16 AM	8:07 AM
21	8:16 AM	9:10 AM
22	8:46 AM	9:24 AM
23	7:07 AM	7:46 AM
24	8:25 AM	9:12 AM
25	7:37 AM	8:25 AM
26	9:15 AM	10:06 AM
27	8:42 AM	9:16 AM
28	7:58 AM	8:45 AM
29	7:04 AM	7:51 AM
30	8:30 AM	9:16 AM
31	7:06 AM	7:49 AM
32	9:13 AM	10:19 AM
33	7:18 AM	8:24 AM
34	9:16 AM	10:28 AM
35	8:09 AM	9:16 AM
36	7:40 AM	8:40 AM
37	7:07 AM	8:06 AM
38	8:18 AM	9:07 AM
39	7:07 AM	8:04 AM
40	7:55 AM	8:58 AM
41	7:54 AM	8:46 AM
42	7:30 AM	8:33 AM

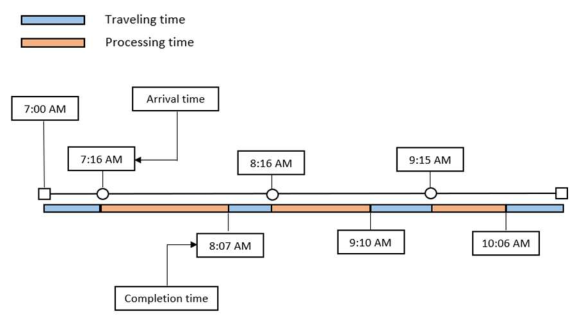


Figure 6: Vehicle 1 (hospital 1) with the visiting sequence: 20→21→26

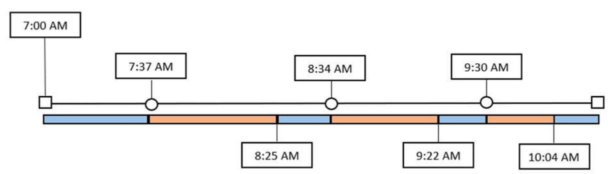


Figure 7: Vehicle 14 (hospital 2) with the visiting sequence: 25 →3→1

The obtained results on both scheduling and routing phases enhance the decision-making process of the hospital in 2 ways:

- As mentioned earlier, hospitals are usually under resource limitations and their finite capability of devoting resources for the home healthcare sector, the solution guarantees that resources can be effectively utilized in the way that all patient demands are fulfilled.
- The generated outcomes of this model also help hospitals to initiate a visiting plan based on the provided patient list for a working shift. The proposed plan provides a time frame for the hospital to make appointments with patients while guarantees that the total completion time of the operation is optimized as well as effectively utilize its limited resources dedicated to the home healthcare sector. Moreover, based on the proposed plan, medical staff can easily manage

and control their time to precisely follow the committed visiting time made with the patients.

5. Conclusion and recommendations

5.1. Conclusion

In this study, the Multi-resources Routing and Scheduling Problem is studied. This paper also formulated a Mixed -Integer Linear Programming model to solve the problem, and a Greedy heuristics algorithm was developed and implemented. The Greedy algorithm was tested over 5 data sets of randomly generated instances with up to 50 patients and 3 hospitals and compared with the exact solution solved by the MILP. The computational results show that the Greedy algorithm can find a better solution within a considerably shorter computational time. It was also tested on real instances related to the case of home healthcare service provided by hospitals in Ho Chi Minh city, where it shows to be effective and able to provide the solution that can both satisfy computational time and optimal values requirements.
This section deals with this paper's contribution based on two key attractions of this problem: the problem aspect complexity and the real-life applications aspect. For the complexity of the problem: two sub-problems are included. The first sub-problem is the resources scheduling problem and the second one is the vehicle routing problem.

- For the resources scheduling problem, many types of resources are considered, they are categorized as expendable and non-expendable resources. The hospital acts as a distribution center that has a finite capacity for each type of resource and also a finite number of vehicles to carry the resources from these distribution centers to patient locations.

- The VRP considered in this research can be categorized as OVRP (Open Vehicle Routing problem) where the vehicles do not require to return to the depot where it dispatched; or if it does, it does not affect the general objectives after all. For the vehicle routing problem (VRP) apart from being a conventional VRP that only considers the shortest paths for the vehicles, this research also extended further by taking the processing time and the weighted factor of every single patient into consideration to make the routing decision.

For the real-life application aspect, the contribution, and benefits that this paper provides are as follows:

- All patients are scheduled to be served by the medical staff and are supplied according to their demand for all types of resources that they need.
- A visiting plan for the hospital for a list of patients guarantees the optimal completion time for each vehicle.
- A complete scheme for all the hospitals under the network for a specific working shift that can be generated quickly helps planners cope with many situations when new patient entry occurs before the planning process takes place.
- An analysis based on the real working schedule of the medical teams provides insights to hospitals to estimate how many more patients can be served apart from the conventional assigning ratio of medical staff per patient.

5.2. Limitations & Recommendations

This paper focuses on addressing the problem of private hospitals in Ho Chi Minh city. The demand for medical services of the citizens is increasing and the private hospitals which have been investing in the modern infrastructure, are aiming at providing services of high quality to the customers. The parties that might benefit from this study are private hospitals in Ho Chi Minh city. For years, many private hospitals have been providing home health care services to their customers. The demand for the service is increasing and to adapt to the change in the market, the planners must acquire a method that supports their operational decision-making process. Stochastic patient entrance is not considered, therefore visiting plan is established at the beginning of every working shift daily.

Future research can deal with long-term planning which can be used to apply to another group of customers in the home healthcare industry which is periodic patient with provided time window as well as the assignment of a specific doctor for a patient is strictly required. Moreover, future studies can extend the problem into stochastic demand, stochastic processing time over the deterministic ones that are studied in this research. Another promising pathway is to develop multi-objective models in which the trade-off between the processing time and weighted factor of patients for selecting the next incident to cover is investigated. Finally, improved heuristics over the Greedy algorithm can be developed to improve the solution which regards to the objective value gap compared with the exact method while still maintaining the advantage of the Greedy algorithm which is short computational time even for large-sized problems.

6. Acknowledgement

The author would like to send gratitude to all of his lecturers at the Industrial Engineering and Management School in International University - Vietnam National University and the Health Care System for Data Collection.

References

Bang, K., Tak, S., Oh, J., Yi, J., Yu, S., & Trung, T. (2017). Health Status and the Demand for Healthcare among the Elderly in the Rural Quoc-Oai District of Hanoi in Vietnam. *Biomed Research International*, 1-13.

National Research Council. (2011). *Health Care Comes Home: The Human Factors*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13149.

Lei, Lei et al. (2015). "Personnel Scheduling and Supplies Provisioning in Emergency Relief Operations." *Annals of Operations Research* 235(1): 487–515.

Pillac, Victor, Pascal Van Hentenryck, and Caroline Even. (2016). "A Conflict-Based Path-Generation Heuristic for Evacuation Planning." *Transportation Research Part B: Methodological* 83: 136–50.

Bodaghi, Behrooz, Ekambaram Palaneeswaran, Shahrooz Shahparvari, and Mahsa Mohammadi. (2020). "Probabilistic Allocation and Scheduling of Multiple Resources for Emergency Operations; a Victorian Bushfire Case Study." *Computers, Environment and Urban Systems* 81(October 2018): 101479.

Bruni, M. E., P. Beraldi, and S. Khodaparasti. (2018). "A Fast Heuristic for Routing in Post-Disaster Humanitarian Relief Logistics." *Transportation Research Procedia* 30: 304–13.

Shi, Yong, Toufik Boudouh, and Olivier Grunder. (2019). "A Robust Optimization for a Home Health Care Routing and Scheduling Problem with Consideration of Uncertain Travel and Service Times." *Transportation Research Part E: Logistics and Transportation Review* 128(November 2018): 52–95.

Zhu, Li, Yeming Gong, Yishui Xu, and Jun Gu. (2019). "Emergency Relief Routing Models for Injured Victims Considering Equity and Priority." *Annals of Operations Research* 283(1–2): 1573–1606.

Lee, Kangbok, Lei Lei, Michael Pinedo, and Shengbin Wang. (2013). "Operations Scheduling with Multiple Resources and Transportation Considerations." *International Journal of Production Research* 51(23–24): 7071–90.

Bodaghi, Behrooz, Shahrooz Shahparvari, et al. (2020). "Multi-Resource Scheduling and Routing for Emergency Recovery Operations." *International Journal of Disaster Risk Reduction* 50(April): 101780.

Moreno, Alfredo, Pedro Munari, and Douglas Alem. (2019). "A Branch-and-Benders-Cut Algorithm for the Crew Scheduling and Routing Problem in Road Restoration." *European Journal of Operational Research* 275(1): 16–34.

Bertazzi, Luca, Leandro C. Coelho, Annarita De Maio, and Demetrio Laganà. (2019). "A Matheuristic Algorithm for the Multi-Depot Inventory Routing Problem." *Transportation Research Part E: Logistics and Transportation Review* 122(January): 524–44.

Ho, William, George T.S. Ho, Ping Ji, and Henry C.W. Lau. (2008). "A Hybrid Genetic Algorithm for the Multi-Depot Vehicle Routing Problem." *Engineering Applications of Artificial Intelligence* 21(4): 548–57.

Wex, Felix, Guido Schryen, Stefan Feuerriegel, and Dirk Neumann. (2014). "Emergency Response in Natural Disaster Management: Allocation and Scheduling of Rescue Units." *European Journal of Operational Research* 235(3): 697–708.

Jung, Kyung Sung, Michael Pinedo, Chelliah Sriskandarajah, and Vikram Tiwari. (2019). "Scheduling Elective Surgeries with Emergency Patients at Shared Operating Rooms." *Production and Operations Management* 28(6): 1407–30.

S. H. Taieb, T. Loukil, and A. El Mhamedi, (2019). "Home (Health)-Care Routing and Scheduling Problem," *Int. Colloq. Logist. Supply Chain Manag. LOGISTIQUA 2019*, pp. 12–14.

A. Majumdar and A. Majumdar, (2019). "Greedy Algorithms," *Compress. Sens. Eng.*, pp. 9–28.

FORECASTING FUTURE PERFORMANCE OF VIETNAM BANKING INDUSTRY USING GREY PREDICTION AND DATA ENVELOPMENT ANALYSIS

Kim-Phong Tran[1*], Chia-Nan Wang[1] and Hector Tibo[1,2]

[1] *Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan, f109142175@nkust.edu.tw, cn.wang@nkust.edu.tw*

[2] *Department of Electrical Engineering, Technological University of the Philippines Taguig, Taguig City 1630, Philippines, hector_tibo@tup.edu.ph*

Abstract. Performance analysis has been applied to many types of industries to measure how efficient the organizations have been doing for a certain period. In this paper, the authors measured the performance of Vietnam's banking industry from the previous years of 2015 to 2019, and future years 2020 to 2023. Using the Grey prediction model, the data for the input factors such as assets, deposits, liabilities, and expenses, as well as the output factors such as loans and interest incomes are calculated. Then applied DEA to the previous and future years to calculate the efficiency index which will represent the performance scores of every bank in Vietnam. Results show that Vietcombank, TPBank, HD Bank, and VietnamMaritime Bank are expected to improve in the future, while VietinBank, SHBank, Oricent Commercial Bank, Saigon Bank, BaoViet Bank, SaigonConthuong Bank and Petrolimex Bank are performing decline. It is recommended for the banks to improve their strategies towards sustainable development as well as improving the lending system to prevent people from getting funding to some informal capital sources.

Keywords: grey prediction; data envelopment analysis; banking industry;

1. Introduction

Performance analysis has been done to many types of industries to measure the capabilities and improve the capacities of the organizations concerned. Many methods were used for performance measurement but in this paper, the method of data envelopment analysis will be used to calculate for the efficiency of the companies that belongs to the banking industry in Vietnam. During the period of 2015 to 2019, most of the Vietnamese banks have shown drop in terms of GDP contribution and asset quality. However to balance the banking boom in Vietnam in the period 2004-2011, the state tightened The State Bank of Vietnam (SBV) tightened monetary policy to fight inflation in 2008-2009. Low economic growth, corporate restructuring, debt leverage reduction, stock market, and real estate plummeting, making credit growth not high in the years 2012-2014 with an average rate of 11% per year. By 2015, credit has increased. In terms of policy orientation, credit growth of 17-18% is considered necessary to ensure GDP growth at 6.6 - 6.8%. Improved corporate profits, business expansion helps increase demand for production loans. But the areas with the highest growth are real estate and consumer credit.

By the end of 2017, concerns emerged among the investor community about macroeconomic stability if credit continues to soar and whether the history of the turbulent cycle will repeat itself. The very positive point is that monetary policy has been more cautious in 2018 and as a result credit only increased by approximately 14%. Meanwhile, the GDP growth rate reached 7.08%, the highest rate in 10 years, and inflation continued to be maintained below the target of 4%. In 2019, the trend is that credit will continue to grow low and if so, in the five years 2015 - 2019, the credit of the whole system will increase by about 2 times. After years of accelerating lending expansion, most commercial banks reported an increase in bad debts since 2011. Although official bad debts according to financial statements only accounted for 4.9% of the total outstanding loans of the entire credit institution system, the actual figure according to the SBV Inspector is up to 17.4%. The timely and effective adjustment has made the banking industry in Vietnam boom and achieved many remarkable achievements during this period. For small and newly established banks in the period 2015-2019, it is a time of challenge as well as potential while balancing the outstanding growth of 15 years ago as well as having to compete in the market. These growth and losses are compiled and presented by an industry analysis report of Ngan (2019).

This paper will use the GM (1,1) grey forecasting model to calculate for the future data to see how the banking industry will perform in the near future and then apply the Window analysis model of DEA to measure their relative efficiency in each two-year periods.

The first section of the paper discusses about a little background of the banking industry, the problem encountered during the period under study, and the goal of applying the grey prediction and DEA Window analysis. Section 2 provides some previous studies that make use of the two models and how they are able to provide useful outputs. Section 3 shows how the models will be use by displaying the framework of the research and discuss the steps involved. Section 4 displays the results from the prediction method and also the efficiency data. There will also be discussion of the data analysis. Lastly, section 5 discourses the findings of the research as concluding statements.

2. Literature Review

The authors have chosen the methods that are believed to be more applicable in the banking industry. Accordingly, the Grey prediction is known to be accurate while using a minimal number of historical data. Some data are really hard to gather especially those that are too old to be organized. And also, DEA is very much applicable to a wide variations of industries as long as data are available. Below are some notable studies which discuss the applications of Grey prediction and DEA.

Xia et al (2015) have proposed a real-time rolling grey forecasting method which will provide a more efficient and acceptably accurate machine health prediction for the novel prognostic. The results of the out-of-sample predictive data imply that there is a noticeable increase in the accuracy of forecast when the proposed method is applied. A short-term power load forecasting method was proposed by Mi et al (2018) which makes use of an improved exponential smoothing grey model in 2018. This method was designed by finding the optimal smooth coefficient aided by the 0.618 golden section method. The research does not just solely improve the reliability and accuracy of short-term power load forecasting but also widens the application scope of Grey prediction and shortens the search interval. Wang et al (2020) used the same prediction model to forecast the data with only historical inputs of six years, from 2013 to 2018. The model was combined with data envelopment analysis to compare the performance of the highly and newly industrialized countries. Liu and Xie (2019) made a variant of the grey prediction by combining the NBGM (1,1) and Weibull cumulative distribution function and compared it to existing variants such as GM (1,1), sole NBGM (1,1), Holt exponential smoothing, and ARIMA. Result of the comparison describes how the new variant WBGM (1,1) becomes more efficient and general than the others. Then a study by Xie and Wang (2017) summarizes how grey prediction evolves within the three decades, how they were applied, and dig into several valuable research direction for grey prediction.

Just like the forecasting method introduced in this chapter, the data envelopment analysis is also applicable in many types of industries and purposes. This method has been known by many researchers and has been applied to many studies. Sueyoshi and Goto (2020) used DEA to measure the performance of some Japanese electric power providers. The environmental assessment aspect of the method was used to investigate how the companies performed right before and after the nuclear power plant disaster in 2011. Nguyen and Tran (2017) was able to match the DEA result coming from predicted values using grey prediction and neural network in the selection of strategic alliance for an Indian firm Calcutta Electric Supply Corporation. Using the combined approach, one company emerged as the best partner in terms of building alliance with the firm. Wang et al. (2019) also made another study with the same approach as the previous but applied it to the airline industry. Sixteen major airlines from Asia were selected to be the decision-making units. Results show that three airline operators turned out to be the best performers in Asia, the Emirates, Sri Lankan and Cebu Pacific. An and Zhai (2020) remodeled DEA and combined with support vector regression for the investigation of the carbon tax pricing in China's thermal power industry. After processing the data, it turned out that there will be a huge demand in the future for China's overall electricity supply wherein the north and the south region shows stronger increase among other provinces

In the past, a few scholarly studies have been conducted to measure and analyze the efficiency performance of the Vietnam banking industry and provided some information whether the industry is doing well or not in other aspects. Minh et al (2013) use the basic DEA models such as the BCC and SBM models for the analysis of the 32 commercial banks in Vietnam for the years 2001 to 2005. In the study, the authors found that being a large commercial bank does not give any guarantee of being highly efficient in comparison to small banks. In addition, state-owned banks have been performing low efficiency which shows that being run by the country's government is not a factor for huge success. A different approach was conducted by Ngan (2014) wherein stochastic frontier analysis (SFA) was used to calculate the cost and profit efficiency of the Vietnamese banking sector for the data 2007 to 2012. In this paper, results show that state-owned banks are more efficient in terms of profit generation than local commercial and foreign-owned banks, while the latter is more cost efficient. Combining SFA with DEA is the goal of a study by Vo and Nguyen (2018) to apply it in analyzing the Vietnam bank industry. The combined models use the data from 1999 to 2015 and applied to 26 commercial banks. After the analysis, the authors concluded that any intervention by the government in terms of policy restructure has negative effect to banks' efficiency. Also, the privatization of some state-owned banks, mergers and acquisition, and any intervention by the government does not make any improvement to their efficiencies. These papers are proof that any efficiency analysis models are applicable to the banking industry of Vietnam and have substantial contributions to the sector if will be used as reference.

3. Methodology

In order to achieve the goals of this study, the following mathematical model and process will guide the authors. The results must be consistent with the purpose of the paper.

3.1. Research Framework

Figure 1 below describes how will the process of research be along with each defining procedures

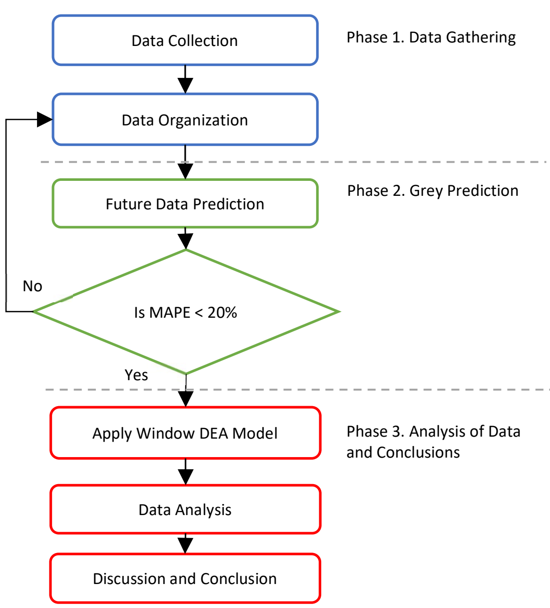


Figure 1. The research process flowchart.

Phase 1. The historical data were gathered from web database of Morningstar.com, a global market research firm. These past data values will be used as the preliminary values (historical data) for the Grey prediction model.

Phase 2. The application of the GM (1,1) is to assist for uncomplicated forecasting by using only a minimum of four years historical data. The forecast will then be used in the process of DEA for evaluation of performance.

Phase 3. All of the data, historical and future, will be processed using DEA Window model to calculate the efficiency during the certain periods. The authors will use the 2-window setup to have a comprehensive comparison between the past and the future.

3.2. Grey Model and MAPE

Based on a time series domain with differential equations in forecasting data, the GM (1,1) has gained its popularity to many users. It became more popular due to its capabilities to produce acceptable and reliable forecast that requires only at least four period of historical data, Ju-long (1982) and Tseng (2001). The process of GM (1,1) is described below:

1. Series x^0 will be encoded as historical data
2. Then x^0 generates the values for $x^1(k)$
3. A partial data will be generated from $x^1(k)$ which is $z^1(k)$
4. The calculation of coefficient a and grey input b
5. Construction of GM (1,1) forecasting equation
6. Calculation of average residual y

Using x^0 as a primitive variable series, the process of the prediction method is described in Equation (1).

$$x^{(0)} = \left[x^{(0)}(1), x^{(0)}(2), \ldots, x(n) \right], n \geq 4 \quad (1)$$

Wherein x^0 is a positive sequence together with the total number of historic data n.

Where n is the total count of historic data along with a positive sequence x^0. A minimum of four historic data is required to facilitate prediction process using grey model. This property is very necessary which makes the model an advantageous method in forecasting.

Equation (2) describes the partial data series:

$$z^{(0)} = [\, z^{(1)}(1), z^{(1)}(2), \ldots, z^{(1)}(n) \,] \quad (2)$$

Where the mean value $z^1(k)$ is also defined in Equation (3) as:

$$Z^{(1)}(k) = \frac{1}{2} \times [X^{(1)}(k) + X^{(1)}(k-1)], \quad k = 2, 3, \ldots, n, \quad (3)$$

The first order differential equation $x^{(1)}(k)$ of grey model can be acquired through Equation (4) as described by Julong (1989):

$$\frac{dx^{(1)}(k)}{dk} + ax^1(k) = b \quad (4)$$

The least square method, Equation (5), will be used to solve the above equation.

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad (5)$$

In which $\hat{x}^{(1)}(k+1)$ represents the predicted value of *x* at *k* +1 point in time. The values of *[a,b]^T* will be generated using the ordinary least square (OLS) method as defined by Equations (6) to (8).

$$[a, b]^T = (B^T B)^{-1} B^T Y \quad (6)$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ x^{(0)}(n) \end{bmatrix} \quad (7)$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \cdots\cdots\cdots & \cdots \\ \cdots\cdots\cdots & \cdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (8)$$

In which $[a, b]^T$ is the parameter series, Y and B are referred to as the data series and data matrix, consecutively.

The $\hat{x}^{(1)}(k)$ values is calculated by having $\hat{x}^{(0)}$ as the predicted series.

$$\hat{x}^{(0)}\, x^{(0)}(1), \hat{x}^{(0)}(2), \ldots, \hat{x}^{(0)}(n) \quad (9)$$

in which $\hat{x}^{(0)}(1)$ is equal to $x^{(0)}(1)$.

The final equation (11) can be obtained by applying the inverse accumulated generation operation (AGO).

$$X^{(0)}(k+1) = \left(X^{(0)}(1) - \frac{b}{a} \right) e^{-ak}(1 - e^{a}) \quad (10)$$

Mean absolute percentage error *(MAPE)* measure the accuracy of the predicted values and is defined by Equation (11).

$$MAPE = \frac{1}{n} \sum \left(\frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right) \times 100\% \quad (11)$$

Figure 1 below describes how will the process of research be along with each defining procedures

The *MAPE* of the predicted data represents its acceptability rate. Wherein lower value depicts highly accurate predictions while higher value means otherwise. Accuracy table was categorized into four by as displayed in Table 1.

Table 1. *MAPE* equivalent forecast categories.

MAPE	**Categorical Value**
10% and below	Highly Accurate
10 to 20%	Moderately Accurate
20 to 50%	Acceptable
51% and higher	Unacceptable

3.3. DEA Window Analysis

In this study, another nonparametric model of DEA will be used and is referred to as the Window model. In this mathematical model, *n* is considered to be the summation of all the units being observed and will be called the DMU_n while the input variable is *m* and output variable is *s*. Integrating an element of time series *t*, it will become DMU_n^t. Then, the input and output generated into a vector called X_n^t and Y_n^t, as describe by the Equations (12) and (13).

$$X_n^t = \begin{bmatrix} x_n^{1t} \\ \vdots \\ x_n^{mt} \end{bmatrix} \quad (12)$$

$$Y_n^t = \begin{bmatrix} y_n^{1t} \\ \vdots \\ y_n^{st} \end{bmatrix} \quad (13)$$

The point $k\ (1 \leq k \leq T)$ is the starting point in time T and has a width magnitude $w\ (1 \leq w \leq T - k)$, wherein each window kw will be represented by the input matrix X_{kw} and output matrix Y_{kw} as shown in the equations below.

$$X_{kw} = \begin{bmatrix} x_1^k & x_2^k & \cdots & x_n^k \\ x_1^{k+1} & x_2^{k+1} & \cdots & x_n^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{k+w} & x_2^{k+w} & \cdots & x_n^{k+w} \end{bmatrix} \quad (14)$$

$$Y_{kw} = \begin{bmatrix} y_1^k & y_2^k & \cdots & y_n^k \\ y_1^{k+1} & y_2^{k+1} & \cdots & y_n^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{k+w} & y_2^{k+w} & \cdots & y_n^{k+w} \end{bmatrix} \quad (15)$$

The DEA Window analysis will commence right after the input and output were substituted to the DMU_n^t equation.

3.4. Selection of DMUs

Another crucial task in the research process is the proper selection of the decision-making units (DMUs). These are the subjects in which the data will be gathered upon and will represent the whole banking industry in Vietnam. So, the authors have selected those banks which have been a contributor not only to the industry but also to the country. Table 2 shows the list of the Vietnamese banks.

Table 2. *List of banks in Vietnam.*

DMU	Bank	DMU	Bank
DMU1	Vietinbank	DMU10	VPBank
DMU2	BIDV	DMU11	NationalCitizen
DMU3	Vietcombank	DMU12	OrientCommercial
DMU4	ACB	DMU13	VietnamMaritime
DMU5	Sacombank	DMU14	SaigonBank
DMU6	SHBank	DMU15	VIB
DMU7	TPBank	DMU16	BaoVietBank
DMU8	HDBank	DMU17	SaigonCongThuong
DMU9	MilitaryBank	DMU18	PetrolimexGroup

With the rapid development of the banking and finance industry, more and more banks are emerging and participating in this potential market with many different functions.

3.5. Identifying Input and Output Factors

After the significant DMUs have been selected, the next step is to identify the input and output factors that will be considered for the analysis. These factors must have important impact to the performance of the banks. The authors have decided to consider those factors that are commonly used by the previous studies. Table 3 lists down the input and output factors and their definitions.

Table 3. *List of banks in Vietnam.*

Input	Description
Total Assets	The total assets are described as the combined assets, values of all the items in that is owned by small businesses.
Deposits from Customers	Customer deposit means that the funds will be credited. It could be money that a company receives from a customer prior to the company earning it
Operating Expenses	Operating expenses are expenses a business incurs in order to keep it running, such as staff wages and office supplies.
Total Liabilities	Total liabilities are the combined debts and obligations that an individual or company owes to outside parties. Everything the company owns is classified as an asset and all amounts the company owes for future obligations are recorded as liabilities.
Output	**Description**
Loans, Advances, and Finance Leases to Customers	A loan is a sum of money that one or more individuals or companies borrow from banks or other financial institutions so as to financially manage planned or unplanned events Customer Advances are reasons a product or service is valuable to a customer. This serves as a guideline for employees who are selling or representing the product to the customer. A finance lease is a type of lease in which a finance company is typically the legal owner of the asset for the duration of the lease, while the lessee not only has operating control over the asset
Net Interest Income	Net interest income is defined as the difference between interest revenues and interest expenses.

4. Data and Results

This chapter discusses the data gathered through the application of the GM (1,1) and the Window model of Data Envelopment Analysis. Using the historical data collected from online sources, these were used to generate predicted data for all the input and output factors. The complete historical and future data were then used for the Window model to generate efficiency index which will be the basis of the banking performance.

4.1. Actual Historical Data

In order to produce future predictions using the grey forecasting model, a required minimum of past or historical data must be collected. These data must conform with the recommended factors that have impact on the performance of the banking industries. The authors have chosen the factors as discussed in the third chapter and summarized the historical data represented by averages in Table 4 below.

Table 4. Average values (in Millions) of the factors for the year 2015 – 2019.

AVE.	2015	2016	2017	2018	2019
TA	237.294	280.744	339.531	372.913	420.029
DC	164.880	203.545	236.361	265.718	302.709
OE	3.475	4.121	5.018	5.699	6.461
TL	217.718	263.844	319.658	350.344	392.468
LFC	144.379	176.712	205.482	242.606	277.788
NII	5.798	6.757	8.665	9.885	12.099

As discussed in the third chapter, GM (1,1) minimum requirement of up to four period of historical data are enough to make accurate forecast. However, the accuracy will depend on the values of historical data if increasing or not. More historical data used will result to more accurate predictions. But also, it must be considered that this study has limitations in the source of data. Since the only available data are from 2015 to 2019, these are enough to be used in GM (1,1).

4.2. Results of Grey Prediction

The grey prediction model was used to project the values of the input and output factors that will be later used for DEA analysis.

Table 5. Average values (in Millions) of the factors for the year 2020 – 2023.

AVE.	2020	2021	2022	2023
TA	546.137	621.804	708.545	808.057
DC	394.106	450.526	515.615	590.822
OE	8.848	10.387	12.249	14.513
TL	508.693	577.827	656.903	747.424
LFC	378.538	442.120	516.959	605.153
NII	17.661	21.506	26.274	32.202

The values of the variables acquired after processing the Grey prediction is noticed to be increasing from 2020 to 2023 as seen in Table X above. The closer the predicted value from the actual values means an accurate prediction. However, the accuracy of the prediction is not measured just in terms of how close the values are to each other. The Mean Absolute Prediction Error formula can calculate how accurate the predictions are. This method was introduced in the third chapter and will now be applied to the acquired date to see if the prediction values are acceptable.

4.3. MAPE Results

Table 6 below shows the average MAPE scores for every DMUs together with respective input and output factors. This is to know whether the predicted values will be accepted before proceeding with the DEA.

Table 6. MAPE results for the year 2020 – 2023.

	TA	DC	OE	TL	LFC	NII
DMU1	1.60%	1.23%	3.00%	1.81%	1.96%	8.01%
DMU2	1.19%	1.02%	1.39%	1.31%	3.83%	5.06%
DMU3	3.28%	0.73%	0.55%	3.58%	0.36%	1.40%
DMU4	0.87%	0.54%	2.93%	1.14%	0.83%	0.97%
DMU5	0.24%	0.90%	1.57%	0.22%	0.58%	2.72%
DMU6	1.58%	0.33%	1.62%	0.52%	1.72%	4.07%
DMU7	1.40%	2.34%	4.67%	1.67%	1.84%	3.03%
DMU8	3.42%	3.41%	1.77%	3.41%	1.35%	1.64%
DMU9	1.51%	0.62%	5.25%	1.78%	0.93%	2.61%
DMU10	0.79%	2.96%	2.93%	0.26%	1.66%	1.83%
DMU11	1.45%	3.11%	5.08%	1.28%	3.53%	4.70%
DMU12	2.23%	1.51%	1.38%	2.46%	1.19%	3.84%
DMU13	1.23%	5.07%	7.34%	1.67%	4.75%	13.00%
DMU14	1.91%	0.70%	4.50%	2.01%	1.63%	15.79%
DMU15	2.86%	4.37%	2.70%	2.94%	1.66%	19.90%
DMU16	5.78%	1.45%	1.66%	6.27%	8.05%	9.45%
DMU17	2.53%	1.07%	2.70%	3.06%	1.99%	1.46%
DMU18	2.22%	2.91%	1.02%	2.56%	2.56%	2.21%

It can be seen above shows the MAPE scores of all the DMUs with respect to each of the variables. It can be seen that most of the scores are in the "Highly Accurate" level which means that the resulting data are acceptable. There are instances of above 10% results for the NII factors of DMU13-15, but these numbers are still in the "Moderately Accurate" level.

4.4. Efficiency Index

Once all data are confirmed to be acceptable, the DEA analysis will proceed. With the use of the DEA Window analysis, the efficiency index of the past in Table 7 and future in Table 8 are calculated.

Table 7. Efficiency indices from 2015 – 2019

	2015	2016	2017	2018	2019	Ave.
DMU1	0.952	0.954	0.989	1.000	1.000	0.979
DMU2	0.998	0.954	0.855	1.000	1.000	0.961
DMU3	0.938	0.810	0.804	0.870	0.897	0.864
DMU4	1.000	0.938	0.939	0.942	0.945	0.953
DMU5	0.846	0.796	0.804	0.838	0.866	0.830
DMU6	0.919	0.960	1.000	0.971	1.000	0.970
DMU7	0.711	0.776	0.806	0.857	0.881	0.806
DMU8	0.726	0.748	0.764	0.829	0.968	0.807
DMU9	0.815	0.829	0.815	0.819	0.846	0.825

DMU10	0.853	0.936	1.000	1.000	1.000	0.958
DMU11	0.569	0.568	0.656	0.706	0.662	0.632
DMU12	0.844	0.826	0.817	0.823	0.885	0.839
DMU13	0.397	0.555	0.539	0.614	0.686	0.558
DMU14	0.920	1.000	0.954	0.922	0.914	0.942
DMU15	0.795	0.864	1.000	0.996	0.958	0.923
DMU16	0.790	0.697	0.842	0.887	0.775	0.798
DMU17	0.937	0.937	0.949	0.954	0.914	0.938
DMU18	0.909	0.980	0.985	0.995	1.000	0.974

Table 8. Efficiency indices from 2019 – 2023

	2019	2020	2021	2022	2023	Ave.
DMU1	0.843	0.879	0.897	0.915	0.934	0.893
DMU2	0.833	0.895	0.929	0.964	1.000	0.924
DMU3	0.736	0.823	0.873	0.926	0.982	0.868
DMU4	0.759	0.772	0.778	0.784	0.792	0.777
DMU5	0.689	0.734	0.763	0.797	0.836	0.764
DMU6	0.845	0.854	0.862	0.879	0.937	0.875
DMU7	0.678	0.799	0.862	0.929	1.000	0.854
DMU8	0.746	0.827	0.882	0.939	1.000	0.879
DMU9	0.708	0.731	0.747	0.776	0.827	0.758
DMU10	0.866	0.924	0.949	0.974	1.000	0.943
DMU11	0.536	0.624	0.672	0.725	0.783	0.668
DMU12	0.697	0.704	0.715	0.727	0.747	0.718
DMU13	0.510	0.603	0.671	0.746	0.830	0.672
DMU14	0.735	0.679	0.672	0.664	0.655	0.681
DMU15	0.794	0.881	0.920	0.960	1.000	0.911
DMU16	0.614	0.662	0.665	0.668	0.670	0.656
DMU17	0.761	0.754	0.744	0.734	0.725	0.744
DMU18	0.847	0.859	0.867	0.876	0.886	0.867

Table 7 and 8 shows that the past periods have more perfect efficiency results as compared to the future periods. It can be noted that the DMU10 has reached high efficiency performance for three consecutive years of 2017 to 2019. However, it was not able to maintain this result as the efficiency of the future data only has a perfect efficiency on the year 2023. DMU1 and 2 both have consecutive 1.000 scores for the years 2018 and 2019, but only DMU2 has again received another perfect efficiency for the future period. Most of the DMUs do not get a score of 1.000 and some have very low efficiencies. Even though DMU7 and 8 did not get any perfect efficiency for most of the past and future period, they are able to score 1.000 for the year 2023 which is resulted from their increasing efficiency.

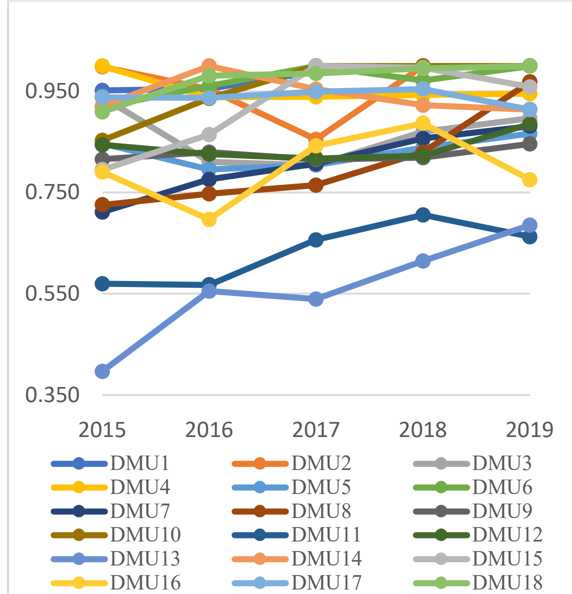


Figure 2. Graphical presentation of the efficiency indices from 2015 to 2019.

Figure 2 above shows that many DMUs are performing in a high efficiency level which is above 0.750. They may not be consistent in every year periods, but it can be noted that they are still high enough. However, two DMUs are performing below this level. DMU11 and 13 perform lower than the others around 0.350 to 0.650 level which is kind of low for performance indices. Also, it can be observed that DMU7 and 8 started in a low efficiency in 2015 but was able to increase by the next periods until they reach 0.890 and 0.950, respectively.

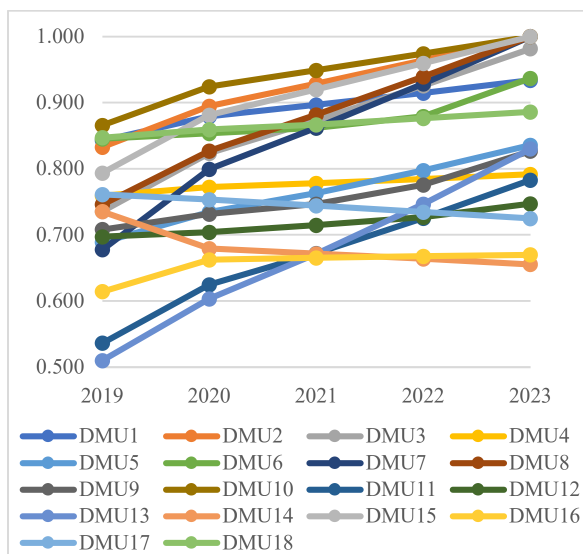


Figure 3. Graphical presentation of the efficiency indices from 2019 to 2023.

Figure 3 shows that DMU2, 7, 8, 10, and 15 are all started in a low efficiency level but was able to reach the highest efficiency of 1.000 on the year 2023 and that show us that these banks are strongly developing throughout the years.

5. Conclusion

Based on the analytical processing method and the results shown in the chart above, we can predict that the upcoming development of Vietnam's banking industry will have many fluctuations and diverse growth. Specifically, we can see that most of the banking situation continues to operate and develops well but not significantly, Banks that will slow down in the future include Vietinbank, SHbank, Oricent Comercial Bank, SaigonBank, BaoVietBank, SaigonCongthuong Bank, Petrolimex Bank. And banks that are expected to see good growth in the future include Vietcombank, TPBank, HD Bank, and VietnamMaritime bank. To explain this situation, we rely on the actual economic situation in recent years.

Firstly, the USD / VND exchange rate policy is stable: Exchange rates of currencies in the world fluctuated strongly in the past year due to the multi-dimensional impact, of which the biggest impact was from the US - China trade war.

Secondly, change in limits and ratios to ensure the safety of the bank's operations. The State Bank of Vietnam has officially issued Circular No. 22/2019 / TT-NHNN replacing Circular No. 36/2014 on safety limits and ratios for banks. This circular has 3 notable points: "tightening" the ratio of short-term capital for medium and long-term loans, increasing the risk ratio for large-value home loans, and adjusting the LDR ratio. Accordingly, banks will have to bring the ratio of short-term capital for medium and long-term loans to 37% from October 1, 2020, continue to decrease to 34% a year later and to 30% after one year.

This regulation will work towards reducing the capital adequacy ratio (CAR) of banks with large home loan portfolios, and at the same time creating opportunities for banks with high CARs to gain market share in home loans. of banks with low CARs, as the lower the CAR, the less room for large-value home loans.

One of the other important reasons is that as of the end of December 2019, the system of commercial banks in Vietnam had 18 members, which were confirmed by the State Bank to apply Circular 41 on Basel II standards ahead of time. Out of these 18 banks, only one bank, VIB, has completed all 3 pillars of Basel II. Meanwhile, there are 2 large members such as Agribank and VietinBank, which still meet Basel II standards.

The next reason is that "Big 4" banks cannot raise capital from the budget. The requirement to raise capital for the "Big 4" banks with dominant state capital has not been able to remove the mechanism and resource arrangement in 2019. Agribank and VietinBank continue to face difficulties in this issue, and BIDV has dissolved by itself by selling capital to KEB Hana Bank; Vietcombank continues to promote plans to sell more capital to foreign investors to increase capital. CAR of the four state-owned commercial banks was close to the allowable threshold. If these banks cannot increase their charter capital, they may have to limit credit extension, even stop extending credit. Currently "Big 4" accounts for about 48% of the credit market. The next reason is that banks have finished settling bad debts at VAMC: By the end of 2019, the banking system had 9 commercial banks that have finished settling bad debts and sold them to Vietnam Asset Management Company (VAMC).

Finally, banks Deploy solutions to limit black credit in provinces and cities to coordinate with the banking industry to promote effective implementation; directing departments, branches, authorities at all levels, especially commune, ward, village and village authorities to coordinate with the banking industry in assisting people to access credit programs for production and business. Restricting people from looking to informal sources of capital.

In contrast to previous studies conducted by several scholars, this paper provided a forecast of data from 2019 to 2023 in which through the Grey prediction. By the accuracy of this forecasting model, and its reliability measure through MAPE, the future data generated from the historical data is used to analyze the future performance of the Vietnamese banks wherein the previous studies are lacking and only focused on the available data. In addition, this paper has provided a more specific standpoint for some banks and shows a bigger picture of what to expect from these banks in the future and how to deal with the uncertainties.

Based on the calculations and simulations using DEA model, we can predict the development of the Vietnamese Banking industry in the coming years, this is an important step to minimize the risks can be occur and at the same time help anticipate opportunities and challenges to be able to build a sustainable development strategy for banks in Vietnam in this financial market. Thereby, banks that have been developing should try to maintain and develop better, and at the same time new banks should have appropriate development strategies and further improve their operations and services to be able to towards sustainability in the future.

References

An, Y., & Zhai, X. (2020). SVR-DEA model of carbon tax pricing for China's thermal power industry. *Science of The Total Environment*, 734, 139438.

Liu, X., & Xie, N. (2019). A nonlinear grey forecasting model with double shape parameters and its application. *Applied Mathematics and Computation*, 360, 203-212.

Mi, J., Fan, L., Duan, X., & Qiu, Y. (2018). Short-term power load forecasting method based on improved exponential smoothing grey model. *Mathematical Problems in Engineering*.

Minh, N. K., Long, G. T., & Hung, N. V. (2013). Efficiency and super-efficiency of commercial banks in Vietnam: Performances and determinants. *Asia-Pacific Journal of Operational Research*, 30(01), 1250047.

Ngan, T. T. (2014). Profit and cost efficiency analysis in banking sector: A case of stochastic frontier approach for Vietnam.

Ngan, T. (2019). Banking Industry Outlook: The Race for Basel II Compliance. *Mirae Asset Daewoo Research*, Jefferies Hong Kong Limited.

Nguyen, N. T., & Tran, T. T. (2017). A novel integration of DEA, GM (1, 1) and neural network in strategic alliance for the Indian electricity organizations. *Journal of Grey System*, 29(2), 80-101.

Sueyoshi, T., & Goto, M. (2020). Performance assessment of Japanese electric power industry: DEA measurement with future impreciseness. *Energies*, 13(2), 490.

Vo, X. V., & Nguyen, H. H. (2018). Bank restructuring and bank efficiency—The case of Vietnam. *Cogent Economics & Finance*, 6(1), 1520423.

Wang, C. N., Tibo, H., & Duong, D. H. (2020). Renewable Energy Utilization Analysis of Highly and Newly Industrialized Countries Using an Undesirable Output Model. *Energies*, 13(10), 2629.

Wang, C. N., Tsai, T. T., Hsu, H. P., & Nguyen, L. H. (2019). Performance evaluation of major asian airline companies using DEA window model and grey theory. *Sustainability*, 11(9), 2701.

Xia, T., Jin, X., Xi, L., Zhang, Y., & Ni, J. (2015). Operating load based real-time rolling grey forecasting for machine health prognosis in dynamic maintenance schedule. *Journal of Intelligent Manufacturing*, 26(2), 269-280.

Xie, N., & Wang, R. (2017). A historic Review of Grey Forecasting Models. *Journal of grey System*, 29(4).

APPLICATION OF MACHINE LEARNING IN MULTI-CLASS CLASSIFICATION: A CASE STUDY OF LIFE INSURANCE RISK ASSESSMENT

Khanh Ho[1] and Son Dao[2]

[1] School of Industrial & Engineering Management, International University, Vietnam National University HCMC Ho Chi Minh City, hogiakhanh.2106@gmail.com

[2] School of Industrial & Engineering Management, International University, Vietnam National University HCMC Ho Chi Minh City, dvtson@hcmiu.edu.vn

Abstract. In the life insurance industry, risk assessment plays a crucial role in decision-making processes that are related to customer application and premium determination. With the availability of the increasing historical customer data and advances in data analytics, life insurance companies are facing an urge to shift toward more automated and data-driven approaches in their underwriting process. Therefore, in this study, we proposed a Machine Learning (ML) model to enhance the applicants' risk classification of life insurance firms. It is a framework using Extreme Gradient Boosting (XGB) algorithm together with a wrapper-based feature selection utilizing the Adaptive Binary Particle Swarm Optimization (ABPSO) to cut down on the input attributes and enhance the predictive performance of the model. Moreover, we applied some methods for model improvement before comparing the final results with various conventional ML algorithms including Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Trees (DT). A real-world dataset provided by Prudential has been used to perform this analysis. The computational results show that my approach is highly competitive, which achieves a higher Quadratic Weighted Kappa score (0.646) and fewer features compared to other methods.

Keywords: Extreme Gradient Boosting; Risk Assessment; Life Insurance; Feature Selection; Particle Swarm Optimization.

1. Introduction

One of the main targets of insurance companies is to effectively sell their policies to customers in which the underwriting process is indispensable as insurance providers need to appropriately evaluate their clients and grant policies. The tasks of the underwriting process include extensive information gathering where customers fill in insurance application forms with their personal information such as age, weight, medical history, family background, etc. for further evaluation. Then, the risk profile of each customer is assessed by the underwriter for decision making on whether their application is acceptable or not. The insurance premiums are calculated afterwards. The lack of proper risk assessment practices may lead to inappropriate premium determination, customer dissatisfaction or financial loss. Therefore, this study proposes a more automatic approach utilizing ML conducted on a case study provided by Prudential, a life insurance provider. Specifically, we implement an XGB model training enhanced by a metaheuristic wrapper-based feature selection, which is the ABPSO, to classify the risk levels of the life insurance applicants. To fully demonstrate the performance of the proposed model, 3 conventional ML algorithms are employed for comparison including RF, DT, and KNN. Section II specifies the literature review of this study; next, the methodology and experimental result are introduced in section III and IV, respectively. Finally, the conclusions are made in section V.

2. Related Works

2.1. Predictive models in risk assessment

As the underwriting process have been considered the backbone of insurance companies' pipeline by several researchers, and predictive analytics are proven to have the ability to improve the applicant risk assessment, many studies deployed ML approaches to predict risk values in the life insurance industry.

Si and Ogunnaike (2017) compared some machine learning models including Linear Regression, Support Vector Regression, RF Regression and Feed Forward Neural Network (FFNN) evaluated by the Mean Squared Error (MSE) for predicting the loss target regression value of a given insurance claim. They found that RF Regression gave markedly better performances with the MSE of 0.390.

Analysing Allstate insurance claim dataset, Chen and Guestrin (2016) also built an effective statistical model to predict the repair cost in each car insurance claim.

They concluded that the scalable ML algorithm for tree boosting, XGB, was applicable to solve real-world scalable cases with limited number of resources.

Henckaerts et al. (2020) also applied tree-based ML methods in insurance tariff plans to build a predictive insurance pricing model. The comparison between Generalized Linear Models (GLMs), RT, RF, and Gradient Boosting machine learning algorithms showed that Boosted trees outperformed the typical GLMs.

Fauzan and Murfi (2018) used the Porto Seguro dataset from Kaggle to build a predictive model for identifying whether a certain customer would file a car insurance claim or not. Comparing XGB with an online learning-based method and ensemble learning including AdaBoost, Stochastic Gradient Boosting, and RF, the study showed that XGB obtained the highest accuracy in terms of normalized Gini.

The risk assessment problem provided by Prudential that this study is conducting on is one of the few ordinal multi-class classification problems in the life insurance industry having been researched. Jain et al. (2019), Mustika et al. (2019), and Boodhun and Jayabalan (2018) conducted on this dataset.

Jain et al. (2019) implemented ML algorithms in which 2 ensemble learning models were deployed to compare with the XGB algorithm. The result of the study showed that XGB without dimensionality reduction outperformed the others with the highest QWK value (64.55), the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The XGB with hyper-parameters optimization was also superior compared to DT, RF, and Bayesian ridge models with a QWK score of (60.73) (Mustika et al. 2019). Boodhun et al. (2018) implemented Multiple Linear Regression, Artificial Neural Network, REPTree and RT with attributes being reduced using Correlation-based Feature Selection (CFS) and Principal Components Analysis. Their study showed that with CFS, REPTree outperformed the others with the lowest MAE of 1.526 and RMSE of 2.027.

To sum up, tree-based algorithms, especially XGB, were proven to be applicable and perform well in studies on risk assessment, so this study employs XGB to classify life insurance applicants' risk levels.

2.2. Feature selection

Feature selection is a crucial technique in which a subset of important features is selected while existing irrelevant or redundant ones in the datasets are ignored. This technique is widely used in ML, especially when the number of predictor variables or the scale of the datasets are large, to decrease the model's classification errors and computational cost. According to Gu et al. (2018), if a dataset has m number of attributes, the number of all possible feature subsets is up to 2^m, which may require metaheuristics to solve high-dimensional problems instead of conventional approaches.

In risk assessment, there have been many papers proposing feature selection methods to remove irrelevant and redundant features. Le et al. (2021) used a wrapper-based feature selection method utilizing Grey Wolf Optimization and PSO to enhance the performance of their early diabetes prediction model. Many other metaheuristics have also been introduced in previous works with high-dimensional datasets, including PSO (Tran et al. 2014), Binary Differential Evolution (Apolloni et al. 2016), and hybrid Ant Colony Optimization (Ma et al. 2021).

To sum up, feature selection techniques, especially metaheuristics, have been proven to be important and can improve the models' performance in many previous works. However, they were paid little attention in the previous works that used the Prudential's dataset. We therefore deploy a metaheuristic feature selection approach to fill the gap.

3. Methodology

3.1. Proposed framework

After the dataset is pre-processed, it applies XGB for model training with a metaheuristic feature selection method to find the optimal subset of attributes. Hyper-parameters and thresholds tuning are also considered for model enhancement. If the result of the model is efficient, it is compared with the results from other conventional ML algorithms. Conclusions are made at the end of the process. Fig. 1 demonstrates the proposed framework of this study.

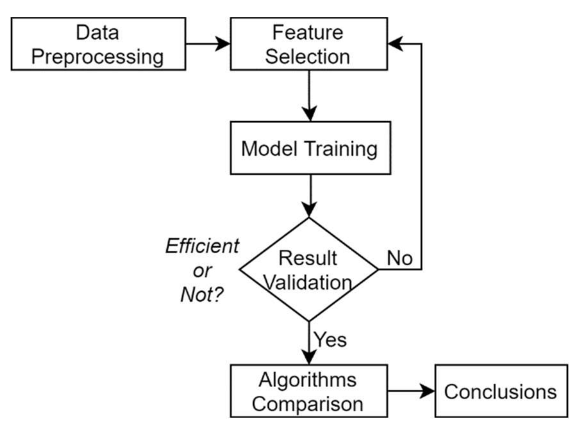


Figure 1. The process flow of the proposed system

3.2. Data pre-processing

The Prudential's dataset includes personal information from 59,381 life insurance applicants with 127 attributes and a response variable corresponding to the risk level of each customer ranging from 1 to 8.

In the first stage of the simulation process, the dataset is split into two parts with a ratio of 80% and 20% for training and evaluating purposes, respectively. According to the plot of the target variable distribution in Fig. 2., the degree of class imbalance is quite significant. Therefore, to avoid poor predictive performance, specifically for the minority classes, the dataset is split proportionally based on the class distribution using the stratify parameter in the scikit-learn library.

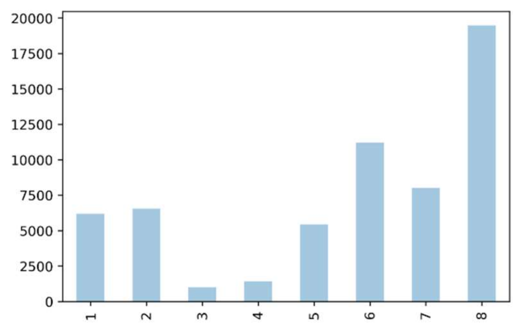


Figure 2. Distribution of the "Response" variable

Due to the nature of the life insurance industry, in the applicant's historical data, there will be the high possibility of missing values when customers fill in the application forms, which is necessary to be dealt with to obtain better performance in the risk assessment process. Table 1. illustrates the percentage of all missing values in the dataset.

Table 1. The percentage of missing values

Features	Missing percentage
Medical_History_10	99.091 %
Medical_History_32	98.135 %
Medical_History_24	93.542 %
Medical_History_15	75.067 %
Family_Hist_5	70.424 %
Family_Hist_3	57.561 %
Family_Hist_2	48.280 %
Insurance_History_5	42.636 %
Family_Hist_4	32.275 %
Employment_Info_6	18.238 %
Medical_History_1	14.929 %
Employment_Info_4	11.403 %

The 3 first medical history features in the table have a high percentage of missing values of up to above 90%. However, assumptions about the importance of these features still could not be made as their missing values can be introduced due to various reasons. As XGB can handle missing data, replacing missing values with -1 is the applied method.

The categorical variable "Product_Info_2" is under text form. Therefore, Label Encoding is the chosen approach for converting this feature into numeric, which is illustrated in Fig. 3. as follow:

D3		D3	D	3		0	0	0
A1		A1	A	1		1	1	1
E1	→	E1	E	1	→	2	2	1
D3		D3	D	3		0	0	2
D2		D2	D	2		4	0	3

Figure 3. Illustration of label encoding

For the other categorical variables in the dataset, Frequency Encoding is applied.

After being standardized, the dataset is put into Pearson's correlation coefficient analysis. This study uses heatmap to perform the analysis, but due to the huge size of the computed map, it is not demonstrated in this study. We follow an example of a conventional approach to interpreting the correlation coefficient, which is illustrated in Table 2. (Schober et al. 2018).

Table 2. A conventional approach to interpret a correlation coefficient

Absolute Magnitude	Interpretation
0.00 – 0.10	Negligible
0.10 – 0.39	Week
0.40 – 0.69	Moderate
0.70 – 0.89	Strong
0.90 – 1.00	Very Strong

From the Pearson's correlation heatmap, 10 pairs of features with "Very Strong" correlation are identified. As this study eliminates only "Very Strong" relationships, 10 features are decided to be dropped after testing, which are "Insurance_History_3", "Insurance_History_4", "Insurance_History_7", "Insurance_History_9", "Medical_History_37", "Medical_History_25", "Medical_History_36", "Medical_History_6", "Medical_History_33", and "Family_Hist_4". After the elimination, 118 features are left for prediction excluding the target label and "ID" feature.

3.3. Supervised ML algorithms

3.3.1. XGB

XGB is an ensemble ML algorithm based on DT, which is developed from a gradient boosting framework that delivers superior performance in terms of both accuracy and computational time. In XGB, the decision to use threshold is made without worrying about how leaves are split later thanks to a split determination algorithm called the Exact Greedy Algorithm. Moreover, another algorithm that helps XGB to deliver superior performance is the sparsity-aware split finding, which is a part of the CART generation process that enables missing values in the dataset to be accurately and internally handled.

The algorithm tries to minimize the following function:

$$O = \sum_{i=1}^{n} L(y_i, \hat{y_i}) + \sum_{k=1}^{N} \Omega(f_k) \quad (1)$$

where n is the number of observations, y_i is the i^{th} observation value of the dataset, $\hat{y_i}$ is the i^{th} prediction value of the target variable, N is the number of decision trees, and f_k is the k^{th} prediction function.

L is denoted as the loss function representing the fitness function of the model and Ω is the regularization term that controls the model's overfitting by managing its complexity. The loss function can be approximated using the Second-Order Taylor Approximation. The regularization term uses Ridge Regression and performs pruning as follow:

$$\Omega(f) = \frac{1}{2}\lambda \sum_{j=1}^{T} {\omega_j}^2 + \gamma T \quad (2)$$

where λ is the regularization parameter, γ is the penalty parameter, T is the total number of terminal nodes, and $\sum_{j=1}^{T} {\omega_j}^2$ is the Euclidean norm of leaf node j scores that controls the model's complexity. As γT is a constant term, it plays no role in deriving the optimal output values. Therefore, it can be omitted in the objective function. After using the Second-Order Taylor Approximation for each loss function, expanding the regularization term Ω and omitting γT, the objective function at t^{th} iteration is defined in (3).

$$O(t) = \sum_{i=1}^{n} [G_i \sum_{j=1}^{T} {\omega_j}^2 + \frac{1}{2}(\lambda + H_i) \sum_{j=1}^{T} {\omega_j}^2] \quad (3)$$

in which G_i and H_i represent the first and second derivative of the loss function, respectively.

3.3.2. DT

DT is a simple supervised ML algorithm whose structure contains a root node at the top of the tree, internal nodes as the model's attributes with branches representing their possible decision values, and leaf nodes corresponding to the final outcomes of the decisions. To identify the importance of each separation, the algorithm measures and compares the impurity by calculating the Gini index based on the probability of each class as follow:

$$Gini = 1 - \sum_{i=1}^{K} {p_i}^2 \quad (4)$$

with K being the total number of class and p_i being the probability of class i. The node that has the smallest Gini value is assigned to be the root node of the tree. DT can automatically select important features from the dataset. However, this may often cause overfitting to the model.

3.3.3. RF

RF is a supervised ML algorithm based on the ensemble learning concept in which multiple RF classifiers are combined, so it has both the simplicity and flexibility to solve complex problems and improve the model's accuracy. Therefore, RF is well-known for its ability to classify practical datasets with relatively high accuracy and low computational time. However, its main drawback is that it is hard to interpret the model in a comprehensible form, which often causes the RF to be considered as a black-box one (Song et al. 2013). The below figure shows the conceptual diagram of RF.

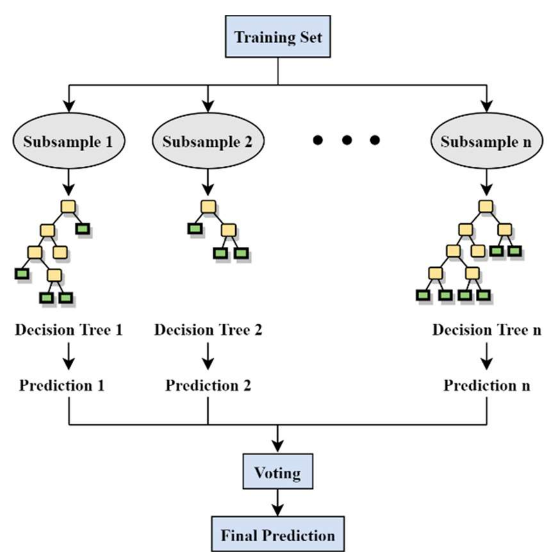


Figure 4. Conceptual diagram of RF algorithm

3.3.4. KNN

KNN is one of the first non-parametric methods, which is commonly used as a supervised ML algorithm for classification problems. After the nearest neighbours is predetermined, to classify an unknown datapoint, the algorithm measures the distance between that point and every other Query point in the training dataset usually by calculating the Euclidean distance (Deng et al. 2016). Then, these distances are sorted into ascending order and the K closest points are returned. The algorithm assigns class to the testing Query point based on the most frequent class of the nearest neighbours. According to Deng et. al (2016), for large-scale datasets, the performance of KNN can be computationally complex and expensive.

3.4. Evaluation metric

Kappa index is a popular statistical coefficient first introduced by Cohen (1960), which considered the inter-judge agreement between 2 judges in nominal scaling. The QWK considers the weight of disagreement between 2 judges defined as follow:

$$\kappa = 1 - \frac{\sum_{i,j} \omega_{i,j} O_{i,j}}{\sum_{i,j} \omega_{i,j} E_{i,j}} \qquad (5)$$

where:
$i, j \in \{1,2,\ldots,C\}$;
C is the total number of classification classes;
$O_{i,j}$ is a matrix corresponding to the number of observations that have the rating of i^{th} and j^{th} categories, which represent the predicted and actual values, respectively;
$E_{i,j}$ is a histogram matrix that calculates the outer product of the predicted and actual classification histogram vectors with all values being normalized so that the sum of all $E_{i,j}$ is equal to the sum of al $O_{i,j}$;
$\omega_{i,j}$ is the weight penalization that is defined based on the ratio $\frac{(i-j)^2}{(C-1)^2}$.

3.5. A wrapper-based feature selection

For classification problems with high-dimensional dataset, swarm intelligence algorithms are considered to have higher performance that can enhance learning models compared to many other feature selection methods (Saw and Hnin Myint 2019). Dara and Banka (2014), Zhang et al. (2014), and Chuang et al. (2011) proposed PSO algorithms to select optimal subset of features from their high-dimensional datasets and showed effective results. Therefore, this study employs a metaheuristic wrapper-based feature selection utilizing the ABPSO to reduce the input attributes and enhance the performance of the model.

To apply the algorithm for feature selection, a fitness function must be initialized. Because the target variable is rendered as an ordinal variable, we choose linear regression as the objective for XGB. Accordingly, the default loss function of XGB is RMSE. After training XGB with this loss function, an error rate is calculated based on the QWK, and (6) is the fitness function that this study tries to minimize.

$$Minimize\ E \times \alpha + (1 - \alpha) \times \frac{N_S}{N_T} \qquad (6)$$

where α, which is set to be 0.95, is a hyper-parameter that decides the trade-off between the classifier performance and the size of the feature subset; N_S is the number of selected features; N_T is the number of total features of the original input data, which is 118.

To find the optimal solution for (6), each particle in PSO stores two values in memory that are continuously updated in each iteration, called the present best position P_i and global best position P_g with the best fitness values of p_{best} and g_{best}, respectively. The position of the i^{th} particle is denoted as X_i in which $X_i = (X_{i1}, X_2, X_{i3}, \ldots, X_{id}, \ldots, X_{in})$. The velocity of the i^{th} particle is denoted as V_i in which $V_i = (V_{i1}, V_2, V_{i3}, \ldots, V_{id}, \ldots, V_{in})$. In each iteration of search for the optimal solution, the algorithm updates each particle's position and velocity as follows:

$$V_{id}^{t+1} = \omega V_{id}^{t} + c_1 \rho_1 (P_{id}^{t} - X_{id}^{t}) + c_2 \rho_2 \left(P_g^t - X_{id}^t\right) \qquad (7)$$

$$X_{id}^{t+1} = X_{id}^{t} + V_{id}^{t+1} \qquad (8)$$

where $d = 1, 2, \ldots, n$ is each particle's dimension; ω is the inertia weight responsible for balancing the global and local exploration capability that helps the algorithm to sufficiently gain the optimal solution; the cognitive acceleration coefficient c_1 and the social acceleration coefficient c_2 are positive constant parameters that impact on the weighting factors of the stochastic acceleration and bring each particle closer to the present best and global best positions; both ρ1 and ρ2 are random numbers ranging from 0 to 1; $t = 1, 2, \ldots, T$ with T being the total number of iterations. The algorithm keeps on performing this process until the g_{best} value converges, or the optimal solution of the fitness function is reached.

To solve a feature selection problem, the real continuous velocity values need to be mapped into probability values with interval [0,1] through a binary mechanism utilizing the Sigmoid function as follow:

$$V_{id}^{t\,\prime} = S(V_{id}^{t}) = \frac{1}{1 + e^{-V_{id}^{t}}} \qquad (9)$$

in which $S(V_{id}^t)$ is the probability value for a bit of a particle changing to 1. V_{id} is assigned to range from a lower bound V_{min} to an upper bound V_{max} to avoid the value of $S(V_{id}^t)$ being too close to 0 or 1. The new position of the particle is updated with (10):

$$X_{id}^{t+1} = \begin{cases} 1, & if\ \rho < S(V_{id}^{t+1}) \\ 0, & otherwise \end{cases} \quad (10)$$

in which ρ is a random number in the range [0,1]; X_{id}^{t+1} is the binary value for the feature selection problem.

Managing the values of the cognitive and social acceleration coefficient is necessary to efficiently find the optimal solution. Therefore, to efficiently balance the algorithm's exploitation and exploration capability, we follow a method proposed by Ratnaweera et al. (2004) and proven to perform well by Ghanem et al. (2018) and Aote et al. (2016), which is to linearly decrease and increase the cognitive and social acceleration coefficient over time, respectively. Their values are defined in (11) and (12).

$$c_1^t = 2.5 - 2.0 \times \frac{t}{T} \quad (11)$$

$$c_2^t = 0.5 + 2.0 \times \frac{t}{T} \quad (12)$$

Inertia weight is also a crucial factor in balancing the exploitation and exploration capability as it can identify the algorithm's searching scope and control the weight of the previous velocity's impact on the new one. Although the linear inertia weight variation is often used in previous works, non-linear one is superior in fitting and simulation, so an inertia weight with logistic chaos is proposed by Liu et al. (2020):

$$r^0 = \rho \notin \{0, 0.25, 0.5, 0.75, 1\} \quad (13)$$

$$r^{t+1} = 4r^t(1 - r^t) \quad (14)$$

$$\omega^t = r^t\omega_{min} + \frac{(\omega_{max} - \omega_{min})t}{T} \quad (15)$$

where ρ is a random number in the range [0,1]; the value of ω_{min} and ω_{max} is 0.4 and 0.9, respectively. The population size and maximum runs are set to be 30 and 5,000, respectively. These 5,000 iterations are run 5 times. The final result is the best fitness value found after these runs. Fig. 5. shows the variation of the inertia weight over time.

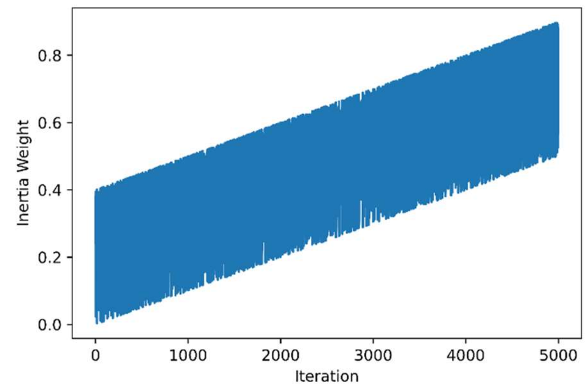


Figure 5. The variation of the inertia weight over time

The pseudo-code of the Adaptive BPSO algorithm is shown in the below figure.

Algorithm 1 The ABPSO Algorithm

Input: V_{min}, V_{max}, *MaxIter, Population*
Output: Array with binary values corresponding to each selection decision
Start: Initialize with random population, binary particles' position and velocity
while (MaxIter) **do**
  for (i = 1 to Population) **do**
    Evaluate the fitness value $f(x_i)$ (6)
    if $(f(x_i) > f(p_{best_i}))$ **then**
      $x_i = p_{best_i}$
    end if
    if $(f(x_i) > f(g_{best_i}))$ **then**
      $x_i = g_{best_i}$
    end if
    for (d = 1 to PopulationSize) **do**
      Update c_1 and c_2 (11) (12)
      Update ω (13) (14) (15)
      Update velocity vectors (7)
      if $V_{id}^{t+1} > V_{max}$ **then**
        $V_{id}^{t+1} = V_{max}$
      end if
      if $V_{id}^{t+1} < V_{min}$ **then**
        $V_{id}^{t+1} = V_{min}$
      end if
      Compute Sigmoid function (9)
      Update position vectors (10)
    end for
  end for
end while
End

Figure 6. Pseudo-code of the ABPSO

4. Experimental Result

After the data pre-processing step, the model training using XGB obtains a QWK score of 0.600. The model is then trained with both Binary PSO (BPSO) and ABPSO for performances comparison. Their best results are obtained after 5 runs, each of which contains 5,000 iterations. Both give approximately equal QWK values of 0.603, but the BPSO's number of selected features is 86 while the ABPSO's is only 74. Fig. 6. shows the convergence curves of BPSO and ABPSO in which the latter has faster convergence speed and reaches the smaller fitness value compared to the former. The fitness values of BPSO and ABPSO are 0.414 and 0.409, respectively, showing that ABPSO performs better in minimizing the model's fitness function.

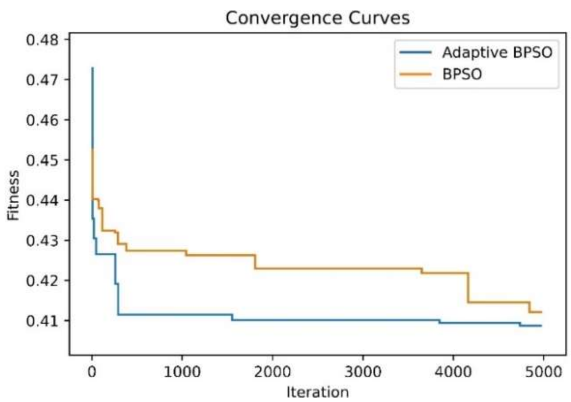


Figure 7. BPSO and ABPSO's convergence curves

Next, the training dataset is split with a ratio of 80% and 20% for training and validation, respectively. After grid searching, the optimized values of "max_depth", "min_child_weight", "subsample", "colsample_bytree", and "eta" are 6, 1, 0.8, 1, and 0.1, respectively. Fig. 8. shows the model's loss after training 500 rounds of iterations. The model has an increasing possibility of overfitting for an increasing number of training rounds, so the early stopping function in XGB is applied to avoid overfitting. The model obtains the best QWK after 487 rounds.

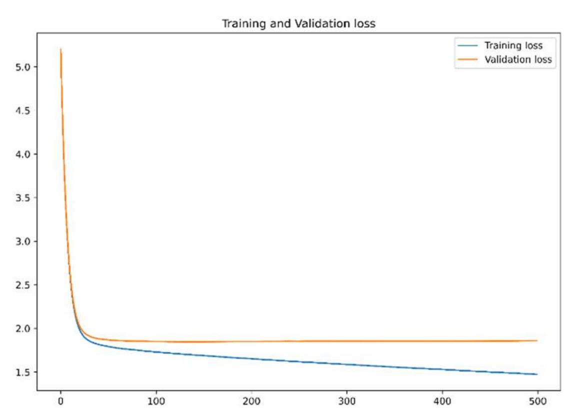


Figure 8. The training and validation loss of the model

The final improvement step is thresholds optimization. The 8 initial thresholds corresponding to the 8 target labels are set to be 0.5. The objective function for thresholds optimization is the model's error rate with predicted values from the training dataset that are added corresponding thresholds. Powell's method is used to minimize this objective function (Powell 1964). After applying hyper-parameters and thresholds tuning, the QWK score of the model increases to 0.646. Fig. 9 shows the confusion matrix of the proposed model.

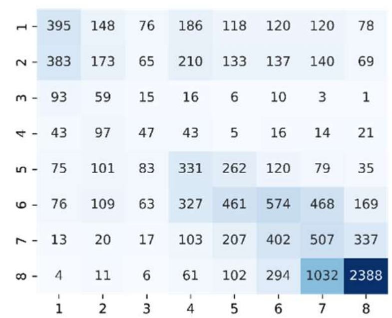


Figure 9. The model's confusion matrix

Table 3. Comparison of the 4 algorithms

Algorithms	QWK
RF	0.508
DT	0.401
KNN	0.220
Optimized XGB model	0.646

Table 3. demonstrates the comparison between the 3 conventional ML algorithms and the optimized XGB model proposed in this study. It can be observed from the table that the optimized XGB model trained with 74 features is superior to the others. According to Landis et. al (1977), with the QWK score of 0.646, the proposed model can be interpreted as substantial based on their measurements of agreement.

5. Conclusions

We have proposed an XGB model training framework together with a metaheuristic wrapper-based feature selection and some model improvement approaches. The proposed model has selected 74 out of the 127 features in the original dataset and obtains the highest QWK score, which is 0.646, compared to the other conventional ML algorithms. Moreover, it is improved in terms of both performance score and dimensionality reduction compared to the aforementioned studies. In future work, we will focus more on enhancing the performance of the ABPSO with stochastic and mainstream learning referring to the study conducted by Liu et al. (2020).

References

Aote, S. S., Raghuwanshi, M. M., & Malik, L. G. (2015). Improved Particle Swarm Optimization Based on Natural Flocking Behavior. *Arabian Journal for Science and Engineering*, **41**(3), 1067–1076.

Apolloni, J., Leguizamón, G., & Alba, E. (2016). Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. *Applied Soft Computing*, **38**, 922–932.

Boodhun, N., & Jayabalan, M. (2018). Risk prediction in life insurance industry using supervised learning algorithms. *Complex & Intelligent Systems*, **4**(2), 145–154.

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Chuang, L. Y., Yang, C. H., & Li, J. C. (2011). Chaotic maps based on binary particle swarm optimization for feature selection. *Applied Soft Computing*, **11**(1), 239–248.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**(1), 37–46.

Dara, S., & Banka, H. (2014). A binary PSO feature selection algorithm for gene expression data. *2014 International Conference on Advances in Communication and Computing Technologies (ICACACT 2014)*.

Deng, Z., Zhu, X., Cheng, D., Zong, M., & Zhang, S. (2016). Efficient k NN classification algorithm for big data. *Neurocomputing*, **195**, 143–148.

Fauzan, M. A., & Murfi, H. (2018). The accuracy of XGBoost for insurance claim prediction. *International Journal of Advances in Soft Computing and its Applications*, **10**(2), 159-171

Ghanem, W. R., Essam, R., & Dessouky, M. (2018). Proposed particle swarm optimization approaches for detection and localization of the primary user emulation attack in cognitive radio networks. *2018 35th National Radio Science Conference (NRSC)*.

Gu, S., Cheng, R., & Jin, Y. (2016). Feature selection for high-dimensional classification using a competitive swarm optimizer. *Soft Computing*, **22**(3), 811–822.

Henckaerts, R., Côté, M. P., Antonio, K., & Verbelen, R. (2020). Boosting Insights in Insurance Tariff Plans with Tree-Based Machine Learning Methods. *North American Actuarial Journal*, **25**(2), 255–285.

Jain, R., Alzubi, J. A., Jain, N., & Joshi, P. (2019). Assessing risk in life insurance using ensemble learning. *Journal of Intelligent & Fuzzy Systems*, **37**(2), 2969–2980.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, **33**(1), 159.

Le, T. M., Vo, T. M., Pham, T. N., & Dao, S. V. T. (2021). A Novel Wrapper–Based Feature Selection for Early Diabetes Prediction Enhanced With a Metaheuristic. *IEEE Access*, **9**, 7869–7884.

Liu, H., Zhang, X. W., & Tu, L. P. (2020). A modified particle swarm optimization using adaptive strategy. *Expert Systems with Applications*, **152**, 113353.

Ma, W., Zhou, X., Zhu, H., Li, L., & Jiao, L. (2021). A two-stage hybrid ant colony optimization for high-dimensional feature selection. *Pattern Recognition*, **116**, 107933.

Mustika, W. F., Murfi, H., & Widyaningsih, Y. (2019). Analysis Accuracy of XGBoost Model for Multiclass Classification - A Case Study of Applicant Level Risk Prediction for Life Insurance. *2019 5th International Conference on Science in Information Technology (ICSITech)*.

Ogunnaike, R. M., & Si, D. (2017). Prediction of Insurance Claim Severity Loss Using Regression Models. *Machine Learning and Data Mining in Pattern Recognition*, 233–247.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, **7**(2), 155–162.

Ratnaweera, A., Halgamuge, S., & Watson, H. (2004). Self-Organizing Hierarchical Particle Swarm Optimizer With Time-Varying Acceleration Coefficients. *IEEE Transactions on Evolutionary Computation*, **8**(3), 240–255.

Saw, T., & Hnin Myint, P. (2019). Swarm Intelligence Based Feature Selection for High Dimensional Classification: A Literature Survey. International Journal of Computer (IJC), **33**(1), 69-83.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients. *Anesthesia & Analgesia*, **126**(5), 1763–1768.

Song, L., Langfelder, P., & Horvath, S. (2013). Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC Bioinformatics*, **14**(1), 5.

Tran, B., Xue, B., & Zhang, M. (2014). Improved PSO for Feature Selection on High-Dimensional Datasets. *Lecture Notes in Computer Science*, 503–515.

Zhang, Y., Wang, S., Phillips, P., & Ji, G. (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowledge-Based Systems*, **64**, 22–31.

MATHEMATICAL OPTIMIZATION MODEL FOR DISPATCHING RIDE-HAILING

Cao Nguyen Dang Minh[1] and Nguyen Vo Cong Thanh[2]

[1]*Product Owner, Be Group, Ho Chi Minh City, caominh1696@gmail.com.*

[2] *Department of Industrial and Systems Engineering, International University, Ho Chi Minh City, nvcthanh1706@gmail.com.*

Abstract. In today's era, technology is the most important factor shaping the general trend of the era due to its versatility and diversity in almost every area of activity in human life. The explosion of technology has led to an explosion of online ride-hailing services. Currently, there are many corporate groups operating in this field with the aim of creating a significant competitive advantage in the market by enhancing the quality of service delivery. In this article, a method covered by optimizing in dispatching ride-hailing through building a mathematical model to analyze the factors affecting user experiences to improve service quality. In the first stage, forming a survey of many customers to determine these factors. Next is the classification, statistics, and analysis of the data source to establish the weights, variables, and constraints needed to build a mathematical model to find the optimal solution for the vehicle routing problem. Apply that mathematical model into reality to collect results and evaluate model quality. The dispatch ride-hailing tools were formed in accordance with the survey data and customer needs to solve certain problems, cost-effectiveness, and minimize the cancellation rate.

Keywords: Factors affecting user experiences; Mathematical model; The trip completion rate.

1. Introduction

A decade with the explosion and rapid change of technology and electronic devices has impacted the behaviour, thinking, and habits of people around the globe. Today, technology is important because it helps to promote social development. Employees in all industries themselves must learn to change and catch up with new trends. It is neither acceptable nor acceptable because this is the general trend of the times. From the past until now, motorbike taxi drivers have to go out to find customers. Customers who want to ride have to find the car themselves and then, the driver sets the price, and the customer also sometimes has to pay the price to avoid being overpriced (Muoi, 2016). There is even a scene of winning passengers between different drivers. However, by applying technology via smartphones such as GrabBike and UberMoto, connecting customers with the nearest driver should be quite quick. As long as you have a motorbike and smartphone, anyone from office workers to students can register to become technology motorbike taxi drivers (Khanh, 2020). Working time is completely free, only when you are free, open the phone to receive guests.

Southeast Asia is an attractive online ride-hailing market, attracting many businesses and technology companies to invest and build the market. Ride-hailing apps seem to be entering the race. The race for market share between car manufacturers is becoming more and more fierce when the competition of domestic car manufacturers begins. Car manufacturers focus on creating advantages by implementing a race to reduce fares on each trip and reward policies to attract drivers and customers. All of them launch promotions, reduce prices together, until they cannot stand the financial pressure anymore, they will have to increase prices and cut bonuses. Good pricing policies are an important factor for ride-hailing apps to attract customers and win the race for ride-hailing market share (Jesus, 2021). But a good price is not enough, because in addition, businesses also need to have more elements of service quality and customer experience (Japutra, 2020).

2. Literature review

The traditional Vehicle Routing Problem (VRP) is set in the context of a delivery company. First of all, the model has one or more warehouses, in each warehouse is a fleet of delivery vehicles and a group of drivers and will move through the road system to reach the delivery points (customers). The required solution problem is a list of routes, S, (one for each vehicle, starting at the depot and ending at the depot itself) such that all customer requirements and conditions are met (Nikita, 2018). Satisfaction, besides, transportation costs need to be minimized. Transport costs can be money, distance, time, ... A road system can be depicted graphically, where the arcs are the roads, and the vertices are the intersections of the roads. Each arc has an associated cost, usually the main cost is the length of time of travel and depends on the type of vehicle. To calculate the cost of each route, identify the cost of travel and the travel time between the intersection and the warehouse by placing intersections and warehouses at vertices, and arcs as paths between them. The cost per arc is the lowest cost between two points on the top road system. From here, it leads to have the complete graph. For each pair of points i and j, there exists an arc (i, j) whose cost is C, which is the lowest cost to move from i to j. The travel time t is the total travel time on the arcs connecting i and j. Sometimes the model cannot meet all customer requirements, and in those cases, the business may reduce some customer needs or not deliver to some customers. This manual method will consume the dispatcher a very long and complicated time every day to plan the route (Be Group, 2020). However, the business can be at risk when a dispatcher is absent and a replacement is needed, making achieving good results with manual routing of routes significantly affected. Other firms will similarly coordinate drivers for customers based on the flight path, with the driver with the shortest route according to the web map being allocated. This coordination in reducing the complexity of the vehicle coordination problem as well as operational expenses, but service quality and customer experience requirements are overlooked.

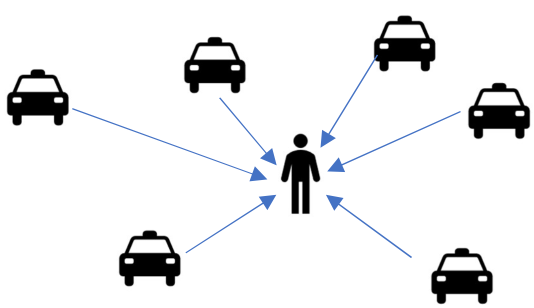


Figure 1: Uber's Grid System

Uber has created a grid system in which points may be contained behind a particular fence, and then the firm evaluates the grid's activity as representing the overall behaviour of the points beneath it as shown in Fig. 1 (Issac, 2018). This implies that each grid may be thought of as reflecting the information contained in the points underneath it. Another problem is that the direct use of points, while useful for humans to see the map and see where the data is gathering, can be computationally a bit more expensive for a system. The computer likely has to calculate the distances of the neighbouring points in order to group them together (See Fig. 2).

Old Way

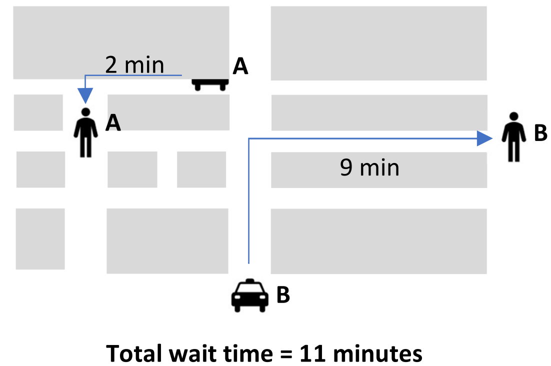


Figure 2: Uber's old method of mapping

Uber has also developed a way to map riders to drivers based on their relationship in the same hexagonal area model. According to the old method, drivers with the shortest distance from customers will be given priority to deliver hoes (Guocheng, 2017). In Uber's new method, instead of combining riders with the nearest driver, Uber's system chooses to coordinate so that the total travel time between neighbouring drivers is the shortest (Ballester-Caudet, 2019). This approach saves the system's operating costs, increases overall customer satisfaction by significantly reducing customer wait times, and reduces the time it takes for drivers to get to the pick-up location that gives them more time to work, and earn extra income (See Fig. 3).

Batched Matching

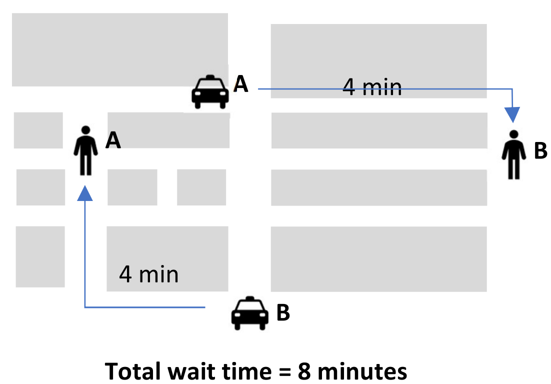


Figure 3: Uber's batched matching method

3. Methodology

A large survey of users of an online car booking application was conducted to find out about demand and factors affecting the experience of using the service. After capturing the influencing factors, we categorize them into smaller groups of factors to highlight the factors that have the most influence on the customer experience. Allocating weights on those factors appropriately is an important thing that helps shape the completion rate of the ride. Experimenting with a pre-existing data set multiple times helps to adjust the weights of the factors reasonably. A mathematical engine has been developed to optimize the dispatch of the most suitable driver to pick up passengers (Grab, 2017). This mathematical tool is based on analysing the relationship between the weights of factors affecting customer experience and customer demand, thereby allocating hoe distribution to drivers with the highest customer satisfaction rate, helping to improve the quality-of-service experience. This mathematical model helps to reduce the operating costs of the system, creating fair competition based on the efficiency of the driver. In particular, helping to create many competitive advantages among car manufacturers in the era of booming technology motorbike taxi service and racing for market share by increasing customer satisfaction when experiencing the service (See Fig. 4).

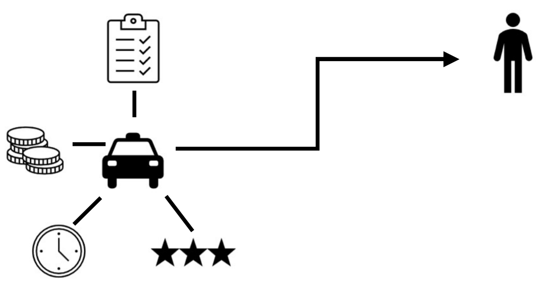


Figure 4: Factors that affect ride-hailing to customers

Factors affecting customer experience are limited and have deviations from reality. The scope of the survey is not large enough and there are not enough factors to quantify. There is insufficient basis to evaluate the validity of the results obtained from the mathematical model, analysing the results based on data collected in the past on the ride-hailing cancellation rate and customer experience. The results of the problem are only true at the time of the survey. Getting the right results for all cases requires a more in-depth approach and allocation.

4. Factor Affecting

4.1. Waiting time for the driver to arrive at the pickup location

There is a request for service to move in the area, based on the map, each driver is a fixed distance from the customer. Waiting is something that no one enjoys. Furthermore, when there is less down time between riders and drivers, drivers may earn more money. The map shows only objective estimates of locations, routes, and distances between the customer and the driver. Each driver has a certain amount of time, from accepting the order to deliver the hoes to picking up customers. Travel time is strictly regulated and controlled by ride-hailing service operators, which is a measure of the driver's ability and working attitude. The waiting time for this activity greatly affects the customer experience, customers do not want to have to wait too long for a designated driver to be served while they can cancel the trip and find their own one. Drivers with shorter waiting times. Furthermore, there are numerous instances where the driver abruptly cancels the trip after leaving the customer waiting for an extended period of time; this has a significant impact on the service quality of corporate firms, and customers frequently boycott the company's services in favour of other ride-hailing services.

4.2. Driver’s star rating level

At the end of the journey, the drivers will receive a rating from customers on a scale of 1 to 5 stars based on their journey experience. The system will average the driver's last 100 or 200 rating points and display the current score on the ride-hailing app. The star rating system helps businesses maintain the service quality standards they always put on top, while also helping to track customer satisfaction and maintain service quality at the highest level. If a driver's star rating falls below the level set by automakers for a long time and shows no sign of improvement after being noted and reminded, the driver's active account will be considered suspended.

4.3. Driver's completion rate

With the desire to bring the best experience to customers and reduce the cancellation rate of drivers, ride-hailing companies always focus on and control the completion rate of each driver. This is a tool to help measure the reliability and responsibility of each driver when working, this indicator will be displayed on the application device every time it is issued. Each driver is encouraged to maintain a certain completion rate to maintain the job and the company's special benefits. Each online ride-hailing company has a different way of calculating the completion rate of each driver. In general, that rate is determined based on the quotient between the rate of Total rides completed and the Total number of trips delivered by the system.

$$Driver's\ completion\ rate = \frac{Total\ rides\ completed}{total\ number\ of\ trips\ delivered\ by\ the\ system}$$

4.4. Driver's income

In order to create fair competition between drivers looking for customers in the same area, it is necessary to consider the income factor to determine the rate of pick-up. Low-income drivers will be given higher priority than drivers with relatively high incomes, this is an incentive for drivers to work hard and help the earning ability of each driver is equal. Currently, technology companies still regularly offer reward programs for drivers. For each level of points the driver receives after each ride, the company will donate the appropriate amount. The development of technology 4.0 has created a large workforce commonly known as technology drivers. In recent years, the number of people doing this profession has become more numerous than ever with all genders and ages. Especially in big cities, the number of motorbike taxi drivers is the largest. According to the survey, the income of a motorbike taxi driver can range from 8 - 15 million/month. In addition to the attractive income, those who work as technology drivers are also active and flexible in terms of working time, if tired, they can rest, without asking anyone's permission.

5. Survey Results

Table 1: Survey information

Ingredient	**Survey Results**		
	Types	**Votes**	**Ratio**
Survey participants	Female	152	0.502
	Male	151	0.498
Age	Under 18	12	0.040
	18 - 24	185	0.611
	25 - 40	42	0.139
	Over 40	64	0.211

Table 2: Factor affecting service quality

Ingredient	**Survey Results**	
	Votes	**Ratio**
Driver's star rating level	129	0.426
Waiting time for the driver to arrive at the pickup location	180	0.594
Driver's attitude	216	0.713
Driver's age	30	0.099
Driver's gender	24	0.079
Driver's uniform	36	0.119

Table 3: Levels of influencing factors affecting service quality

Factor affecting	**Levels**		
	Low	**Medium**	**High**
Driver's star rating level	0.187	0.248	0.565
Waiting time for the driver to arrive at the pickup location	0.175	0.186	0.639
Driver's age	0.564	0.228	0.208
Driver's gender	0.623	0.149	0.228
Driver's income	0.455	0.248	0.297

Table 4: Customer needs based on the survey results

Factor affecting	**Customer needs**		
	Types	**Votes**	**Ratio**
Driver's star rating level	3 – 4	66	0.218
	4 – 5	237	0.782
Waiting time for the driver to arrive at the pickup location	Under 2 minutes	42	0.139
	2 - 5 minutes	156	0.515
	5 - 10 minutes	78	0.257
	10 - 15 minutes	15	0.050
	Over 15 minutes	12	0.040
Driver's age	18 - 24	87	0.287
	24 – 40	195	0.644
	Over 40	21	0.069
Driver's gender	Female	84	0.277
	Male	36	0.119
	Other	27	0.089
	Does not matter gender	156	0.515

According to the survey results in Tab. 1, Tab. 2, Tab. 3, and Tab. 4, our study obtained many types of factors affecting the rider's experience and completed ratio. These factors can indirectly affect the service quality of the technology car company. Types of factors are classified into three main groups: waiting time when the driver moves to pick up passengers, the driver's income per day and driver's ability includes factors such as star rate, attitude, age, gender, uniform, type of vehicle convenience and so on. After classifying the three main factors above, the stage of evaluating the appropriate weight for those three factors is carried out based on the needs of the survey participants. After many tests on the data set in the past to check the reasonableness and fit with reality, we have decided to value the weight of the above three factors according to the table below (Tab. 5):

Table 5: The weight of factor affecting service quality

Factor affecting	Weight value
Waiting time for the driver to arrive at the pickup location	0.65
Driver's performance	0.25
Driver's income	0.1

The next stage is to find a reasonable barrier to the above influencing factors. Considering the customer's needs, we want our mathematical model to meet the appropriate satisfaction limit of the customer when the experience of completing an actual ride with any driver. We conduct surveys and obtain limited levels of customer satisfaction and are consistent with the actual operating system. The results are shown in Tab. 6.

Table 6: Satisfaction level of factor

Factor affecting	Satisfaction level
Waiting time for the driver to arrive at the pickup location	$\leq 15\ mins$
Driver's performance	absolute rate 100%
Driver's income per day	Max 1,000,000

Waiting time for the driver to arrive at the pickup location: Under 15 minutes

The maximum distance from the driver's current location to the rider of the ride-hailing platform in Vietnam is 3 kilometres, meanwhile, customer waiting is 15 minutes (lowest speed is 12km/hr on-peak hours). This data is calculated automatically when the driver arrives at the pickup point.

Driver's performance: Absolute 100%.

The driver's score is calculated from the completed ratio and rating of the driver (1 - 5 stars). The weight of each factor is not mentioned in this paper because it is impacted by more than 20 factors inside. We will apply the sample formula based on this ratio as:

$$Driver's\ performance = 0.8 \times \frac{Ride\ completed}{Request} + 0.2 \times \frac{current\ star\ rate}{5}$$

Driver's income per day: Max 1,000,000

The goal of dispatch is to balance income for all drivers who are online at the same time and same performance. Drivers with a lower income ratio $(total\ income/online\ hour)$ will have a higher probability of getting a ride. 1,000,000 is sample data that is collected from the driver's expectation and benefit of the company. This number will be changed based on the company's situation.

6. Mathematical Model

Parameters

α: the weight of the factor of the waiting time for the driver to arrive at the pickup location

β : the weight of the factor of driver's performance

γ: the weight of the factor of driver's income

Variables

X_{1i}: waiting time for the driver to arrive at the pickup location

X_{2i}: diver's performance

X_{3i}: driver's income

i: indices for driver

S : number of drivers available in the requesting area

Decision Variables

$$Y_i = \begin{cases} 1, if\ driver\ i\ is\ selected \\ 0, if\ driver\ i\ is\ not\ selected \end{cases}$$

P_i: probability to complete the ride

$$P_i = \left(1 - \frac{X_{1i}}{15}\right) \times \alpha + \frac{X_{2i}}{100} \times \beta + \left(1 - \frac{X_{3i}}{1{,}000{,}000}\right) \times \gamma \ (1)$$

Model:

$$Maximize \sum_{i \epsilon S} Y_i P_i$$

Subject to:

$$X_{1i} \le 15$$
$$X_{2i} \le 100$$
$$X_{3i} \le 1{,}000{,}000$$
$$X_{1i} \ge 0$$
$$X_{2i} \ge 0$$
$$X_{3i} \ge 0$$
$$0 \le P_i \le 1$$

7. Computational results and analysis

Our mathematical model is applied to real data from drivers obtained from the past when initiating a ride search at a given location and within a most hexagonal distribution area determination of the operating position of the drivers. We proceed to list the drivers available in the ride search area with their rating factors. Our goal is to find and dispatch the driver with the highest possible ride completion rate among all drivers in the same area based on how we evaluate the weights above.

There are 3 out of 1000 real test cases applied in companies based on historical data giving the following results in Tab. 7, Tab. 8, and Tab. 9:

Table 7: Case 1

i	X_{1i}	X_{2i}	X_{3i}	P_i
1	4	87	779,944	0.7
2	2	88	113,269	0.9
3	5	90	858,728	0.7
4	9	91	23,205	0.6
5	9	100	699,869	0.5
6	10	95	444,934	0.5

Table 8: Case 2

i	X_{1i}	X_{2i}	X_{3i}	P_i
1	10	95	286,843	0.5
2	15	80	964,685	0.2
3	2	99	358,229	0.9
4	11	83	485,256	0.4
5	3	83	832,943	0.7
6	15	92	59,801	0.3
7	4	100	567,675	0.8
8	13	91	972,755	0.3
9	7	94	781,717	0.6
10	13	82	714,226	0.3
11	4	88	390,000	0.8

Table 9: Case 3

i	X_{1i}	X_{2i}	X_{3i}	P_i
1	13	99	509,425	0.4
2	8	86	435,931	0.6
3	6	96	259,765	0.7
4	7	81	347,086	0.6

Applying this formula to 3,400 completed trips in Ho Chi Minh City and Hanoi, the results are as follows (Fig. 5):

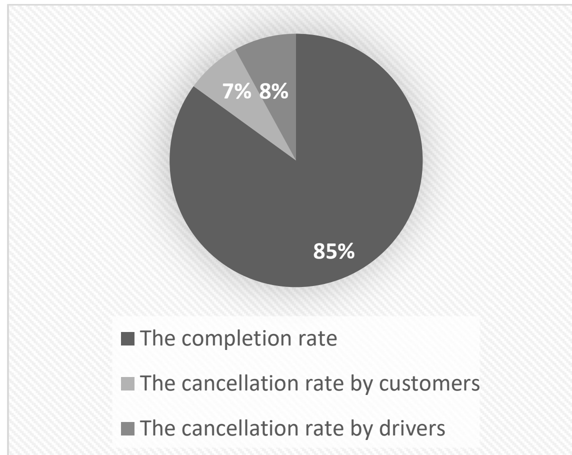


Figure 5: Percentage of cancellation and rides completion

The completion rate of the selected drivers is 85%. The cancellation rate before 30 seconds is 7% by customers because of changes in demand. The remaining 8% were cancelled by the driver, a deep analysis of these drivers found: The amount received for the day was too high, so the lack of motivation and responsibility to complete the trip happened. In the active allocation area, there is a low density of drivers resulting in lower-than-normal driver quality. This is the main reason for the low star rating of the drivers.

8. Conclusions

According to testing results, this model can tackle the majority of situations in hours with full drivers. More than 80% of model selections produce correct outcomes, increasing corporate earnings, and customer satisfaction. This paper concentrates on investigating a mathematical model to help make a multi-criteria decision based on a hierarchical analysis of factors affecting user experience when using an online ride-hailing service based on a wide user survey and how it can be applied in the driver selection process within a defined operating range.

9. Future work

This study, however, has some limitations owing to time restrictions. The factors influencing customer experience are restricted and deviate from reality. The survey's breadth is limited, and there is insufficient evidence to assess the validity of the results. The problem's outcomes are only valid at the time of the survey.

The aim for the next paper will be to develop a mathematical model that can be used for many various types of experience assessment criteria and for many different instances at different periods. The model needs to be improved to be able to handle a variety of time scenarios such as during off-peak hours with no drivers or during periods without many customers. In addition, spatial factors need to be considered such as weather, climate, and topographic factors that affect the behaviour of customers and drivers. A deeper understanding of driver's performance is required to continuously update customer needs and consider assigning appropriate values to the weights. The ultimate goal of this study is to find the right customer with the right driver based on, currently only considering factors and criteria related to drivers, not covering factors from the customer’s side. That will be the work that should be interested in understanding and expanding the problem in the next research paper.

References

Nikita, K., Dawn, W., Chiwei, Y., & Helin, Z. (2018). *Dynamic Pricing and Matching in Ride-Hailing Platforms.*

Muoi, D., & Phuong, M. (2016). *Bùng nổ “xe ôm công nghệ”.* Vietnambiz. https://vietnambiz.vn/bung-no-xe-om-cong-nghe-10989.htm

Be Group. (2020). *Công Thức Tính Tỷ Lệ Nhận Chuyến, Tỷ Lệ Hủy Chuyến, Tỷ Lệ Hoàn Thành.* Be Vietnam. https://be.com.vn/thong-bao-quy-dinh-tai-xe/cong-thuc-tinh-ty-le-moi/

Isaac, B. (2018). *H3: Uber’s Hexagonal Hierarchical Spatial Index.* Uber Engineering. https://eng.uber.com/h3/

Guocheng, X., & Yanjun, H. (2017). *Engineering Uber Predictions in Real Time with ELK.* Uber Engineering. https://eng.uber.com/elk/

Khanh, N. (2020). *Công nghệ ảnh hưởng giới trẻ Việt Nam như thế nào?.* Thanhnien. https://thanhnien.vn/gioi-tre/cong-nghe-anh-huong-gioi-tre-viet-nam-nhu-the-nao-1168491.html

Grab (2017). *Về tính năng đánh giá sao của Grab.* Grab. https://www.grab.com/vn/en/blog/driver/car/ve-tinh-nang-danh-gia-sao-cua-grab/

Japutra, A., Utami, A., F., Molinillo, S., & Ekaputra, I. A. (2020). *Influence of customer application experience and value in use on loyalty toward retailers.*

Ballester-Caudet, A., Campíns-Falcó, P., Pérez, B., Sancho, R., Lorente, M., Sastre, G., & González, C. (2019). *A new tool for evaluating and/or selecting analytical methods: summarizing the information in a hexagon.*

Jesús, C., F., Lily, X., G. Iguácel, M. P. Andreea & T. (2021). *How do firms handle variability in customer experience? A dynamic approach to better understanding customer retention.*

AN IMPROVED MATHEMATICAL MODEL FOR DYNAMIC PRICING IN RIDE-HAILING PLATFORMS

Cao Nguyen Dang Minh[1] and Doan Le Thao Vy[2]

[1] Product Owner, Be Group, caominh1696@gmail.com

[2] School of Industrial Engineering and Management, International University – VNU HCMC, doanlethaovy@gmail.com

Abstract. Ride-hailing platforms, ever since the early days, have undergone massive expansion and transformed public mobility systems. The landscape of ride-sharing promises sustained growth. However, dynamic pricing — a key lever in ride-sharing — presents significant challenges for this industry to overcome to become globally dominant. First, traditional dynamic pricing models only apply demand and supply factors at a certain level when calculating price, which can cause the price to be high and volatile at times, especially during peak hours and holiday season. Besides, wide variation in price between neighboring areas is another downside of current pricing methods leading to a mass of drivers dispatching into areas with higher prices while leaving customers in the lower-price areas with no service. In short, exorbitant price and deficient service result in loss of customer satisfaction and loyalty to the system. To address this issue, we introduce an algorithm that not only computes competitive service prices but also minimizes price deviation for the service. In the final phase of the study, experimental assessments are conducted through authentic data of a major ride-sharing platform to validate our proposed solution and demonstrate its benefits over other fixed pricing schemes.

Keywords: ride-hailing, dynamic pricing, dispatching, experimental assessments

1. Introduction

Ride-hailing refers to the service of booking rides and paying fees through a mobile app. Ride-hailing apps connect passengers and local drivers by the utilization of personal vehicles. The development of this application service by large Transportation Network companies such as Uber or Grab is of great impact on private car use in urban areas (Jun Zhong et al. 2020). It accounts for the effective distribution and deployment of underutilized social resources. In contrast to classical static problems, dynamic pricing strategy continuously makes adjustments to the service price, often in a matter of minutes, in response to real-time demand and supply. Therefore, customers are not able to know the price in advance. In the last decade, dynamic pricing strategies have been proved to productively balance between dynamic supply and demand (Chen and Sheldon, 2015). Its outperformance over static pricing strategies (Vivek Nagraj Pandit et al. 2019) has made it become the trend that reshapes this field of e-commerce industry. According to e-Conomy Southeast Asia 2019 report by Google and Temasek/Bain, Vietnam's ride-hailing sector is expected to reach 4 billion USD by 2025 on a 38% annualized growth rate since 2015.

Dynamic pricing is a powerful mechanism for business's favor that enables businesses to either increase the prices to capitalize on demand or lower prices to promote sales and attract customers. Nevertheless, this pricing scheme is centered on demand and supply factors that are expected to fluctuate widely. For instance, how many people go out on Monday may differ significantly from that on weekends, the number of drivers who are available to provide service may also vary from region to region. With great instability comes great pricing differences.

There are some pain points resulting from this strategy of dynamic pricing. First, high price volatility may affect rider and driver experience negatively. Too high a price can make customers feel cheated and likely to turn away from the platform while too low a price may as well discourage drivers from hitting the road to provide service. Sudden spikes in demand can, on the one hand, be a good chance for business to grow profits and on the other hand, be the cause of service unreliability. As it results in a dramatic increase in the rider waiting time and drop in the capacity utilization rate, leading to a poor experience on both sides. The objective of this study is to develop a dynamic pricing model that balance business welfare with customer satisfaction.

2. Literature Review

Driven by expansive growth of real-life applications, a number of researches have been conducted on pricing strategies in the ride-sharing market. This chapter is focused on a dynamic component to pricing known as surge multiplier and several related work on dynamic pricing strategies.

2.1 Surge multiplier

In 2012, noticing the number of requests that went unfulfilled climbed considerably over weekends leaving many customers dissatisfied, Uber launched the surge pricing model which could come into effect only when there are more demand compared to supply. The purpose of this model is to offer drivers higher prices so that they will stay active on the system for longer periods of time, especially during peak hours (Le Chen at al. 2015). In this system, Uber's ride-sharing marketplace is subdivided into hexagonal hierarchical spatial zones. Fare price in each zone is multiplied by a different surge multiplier that is greater than 1 if demand are higher than supply and equal 1 otherwise.

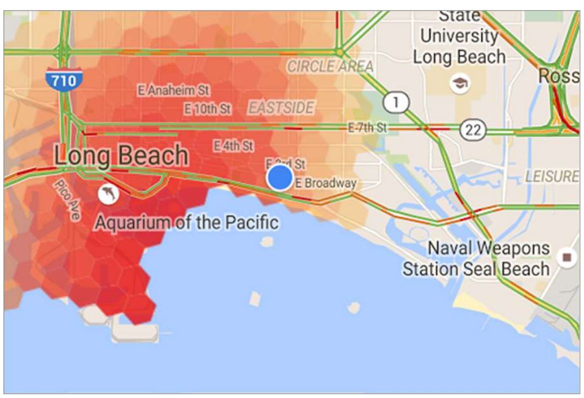


Figure 4: Uber's hexagonal hierarchical network model in Long Beach, California

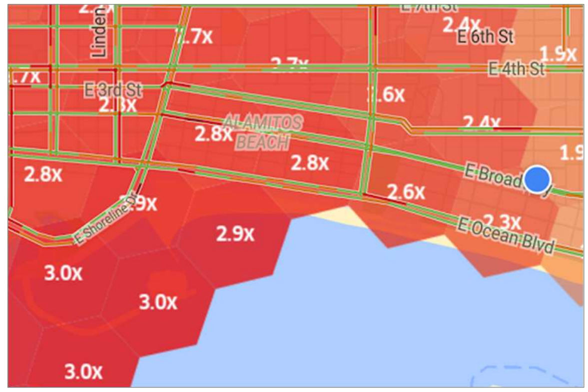


Figure 5: Surge multipliers vary based on location.

2.2 Dynamic pricing

Yan et al. (2019) introduces the strategy of dynamic waiting in which the pool-matching mechanism will flex customer waiting time to mitigate price variation. Two riders whose origin or destination locations are within walkable range will be matched and pooled together in a car. They assume that all riders opt into pooling. Ride waiting time, like price, can also be known through a waiting window before receiving a dispatch. However, in our study, we remain traditional ride booking and serving. Waiting time is excluded from the price model as it is supposed to be investigated in dispatching instead. In a study by Iglesias et al. (2019), the market is a closed, multi-class BCMP queuing network in which customers do not leave the network but only change their locations. Therefore, the number of customers considered demand has a fixed value. Our study, on the other hand, considers demand as the number of customers who make a request by placing an order to the platform only. Since the number of ride bookings can arrive randomly, demand is continuously changing in our model. Bimpikis et al. (2019) explores the spatial price discrimination of ride-sharing market. They emphasize the change in platform's prices, profits, and the induced consumer surplus under the impact of the demand pattern. According to their study, when demand and supply of the whole network are balanced, business achieves maximum profits and consumer surplus at the equilibrium corresponding to the platform's optimal pricing and compensation policy. This becomes their model's objective and their pricing scheme based on spatial dimension will work for each region. However, each routing is priced separately in our model. Another work by Lou and Saigal (2017) proposes a continuous approach to the dynamic pricing problem for on-demand ride-sharing. They study two sides of the market, supply and demand, in a continuous-time continuous-space model to solve the revenue maximization by dynamic programming. This work is helpful for handling the complicated spatiotemporal pricing problem. The improvement in our study is that beside the two most basic lever of ride-sharing problem, supply and demand, we also put in more factors to provide a more realistic solution.

3. Methodology

3.1 Market research and analysis

To acquire an understanding of current ride-sharing market, we carried out a comparative research between three dominant ride-hailing platforms in Viet Nam: Grab, Be, and Gojek (ABI Research, 2019). This research includes a price survey of customer requests that are sent once every fifteen minutes to such platforms, from 7AM to 7PM, as demand in other time ranges is negligible. To obtain a fair result, we only concern base price of each platform which means discount factor is eliminated. Ride bookings are placed on different popular routes in Ho Chi Minh City and prices are recorded on a typical day at weekend. The table below displays our set of price survey data collection.

Table 11: Data collection

Time	Grab price	Be price	Gojek price	Minimum price by	Be usser watch	Be user bookings	% bookings	% drop-off
7:00 AM	31,000	36,000	42,000	Grab	40	8	20	80
8:00 AM								
9:00 AM								
...								
6:30 PM								
6:45 PM								
7:00 PM								

Collected data is analyzed and depicted as price over time in Figure 3 and percentage of Be users' bookings in Figure 4.

Figure 3: Dynamic price of Grab, Be and Gojek

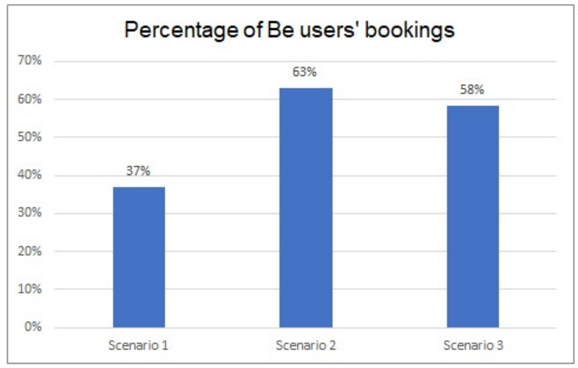


Figure 4: Percentage of Be users' bookings

As shown in the graphs, we can see that price is the most important element in customer decision to experience service. To be specific, when Grab price is checked to be lower than Be price and Gojek price, only 37% of Be users continue to make bookings and complete the trips while the rest 63% of users drop requests and leave platform. Similarly, only 58% of Be users stay loyal while the other 52% will leave for Gojek if Gojek price is the most economical choice.

3.2 Mathematical Model

Our dynamic pricing model is developed based on the principal model of Uber in which we aim to calculate the ideal value of price multiplier. Then, dynamic price is computed as:

$$p = p_b \times M \quad (1)$$

p	dynamic price
p_b	base price at the requested region
M	price multiplier

To determine the value of M, we find the minimum value between *lone surge multiplier* and *regular price multiplier*. This is to ensure that the value of M which the system employs will yield revenue-optimal price without exceeding customer expectation and thus avoid cancellation. Moreover, lower price multipliers increase promotion and encourage customers to experience service when demand is low.

3.2.1 Lone surge multiplier

Lone surge multiplier is a function of demand-supply ratio for current day and time and for last week same day and time. This type of multiplier is highly recommended when demand exceeds supply by a very large amount.

$$V = \begin{cases} \mu_t\lambda_t + \mu_{t'}\lambda_{t'} & \text{if } \mu_t\lambda_t + \mu_{t'}\lambda_{t'} > 1 \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

V lone surge multiplier

μ_t ratio of unique requests to available drivers for current day and time

λ_t weightage of μ_t, $\lambda_t = 0.8$

$\mu_{t'}$ ratio of unique requests to available drivers for last week same day and time

$\lambda_{t'}$ weightage of $\mu_{t'}$, $\lambda_{t'} = 0.2$

3.2.2 Regular price multiplier

Regular price multiplier is drawn from a more complex function of four variables: business benefit, customer drop-off rate, demand-supply nearby ratio, and traffic condition. Set the indices as follow, then we will elaborate on each of four above-mentioned components.

t current index time of the order
i index of price multiplier

Table 12: Price multiplier index

Index (i)	Price multiplier (m_i)
1	1.0
2	1.1
3	1.2
…	…
29	3.8
30	3.9
n = 31	4.0

First, business benefit refers to the maximum benefits our company can obtain upon a given price multiplier and the number of booking rides that the system completes over total customer requests for same day and time last week.

$$b_i = m_i \times \frac{C}{R} \quad (3)$$

b_i business benefit
m_i price multiplier
C/R completed trips/total customer requests

Next, customer drop-off rate is the percentage of customers who leave the platfom without booking upon price multiplier m_i. From the market research which had been conducted and discussed in section 3.1 previously, we observe that during the studied period, price can reach peak at 45,000 VND and hit bottom at 18,000 VND. The highest price difference is correspondingly 2.5 times. An upper limit L of price multiplier above which there are no customer requests is constructed with $L = 2.5 + 0.5(2.5) = 3.7$, then customer drop-off rate d_i (%) is calculated as:

$$d_i = \frac{m_i}{L} \times 100\% \quad (4)$$

Current demand-supply nearby ratio is another indispensable factor to our pricing strategy. Our market research in section 3.1 also indicates that at a given time, not only one platform's supply but the number of available drivers of other platforms also affect customer decision in choosing which service they are willing to pay for. It is recorded that demand is hardly able to exceed supply over 10 times. Let r_t be the ratio of demand to supply nearby at the current time, we have: $r_t \leq 10$.

Lastly, we consider traffic condition p_t when demand pops up (at peak hours or not) a significant factor in our model. The number of vehicles travelling on the road differs from time to time. During peak hours, high traffic volume occurs leading to traffic congestion. This may hinder drivers from arriving at customers' location. Moreover, there is likely to be a surge in demand during this time of the day. As a result, deficient supply can cause unusually very high prices.

$$p_t = \begin{cases} 1 & \text{if current time is at peak hours} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Based on such factors, our goal function is developed to calculate the balance indicator X_i of overall business benefit and customer satisfaction.

$$X_i = \alpha b_i + \beta(1 - d_i) + \omega \times \frac{r_t}{10} + \gamma p_t \quad (6)$$

α weightage of company benefit factor, $\alpha = 0.3$

β weightage of customer drop-off rate factor, $\beta = 0.55$

ω weightage of demand to supply nearby ratio factor, $\omega = 0.1$

γ weightage of current time factor, $\gamma = 0.05$

The higher balance indicator X_i is, the greater company's revenue and better customer experience are. Thus, we aim to find the value m_{i_m}, $i_m \in i = (1$ to $n)$ such that $\sum_{i=1}^{n} X_{i_m}$ is maximum.

In summary, our methodology comprises of two distinct mathematical models that calculate lone surge multiplier and regular price multiplier. To solve for the final result, we need to figure out and compare both types of price multiplier. Its mathematical expression is as below.

$$M = \text{Min}\{V,\ m_{i_m}\} \qquad (7)$$

4. Empirical Analysis

In this section, we first analyze the results obtained from applying the created model into a major real-life ride-hailing platform, then evaluate its practicality. The price model is employed for seven different regions in Ho Chi Minh City and Ha Noi. It is specifically executed during rush hours. The experimental study requires the record and comparison of data from four key business features – customer request, customer retention rate, C/R ratio, and Gross Merchandise Value (GMV) – before and after new pricing policy. Effects of the new model on each aforementioned factor are represented respectively via the charts in Figure 5, 6, 7, 8 and further explained below.

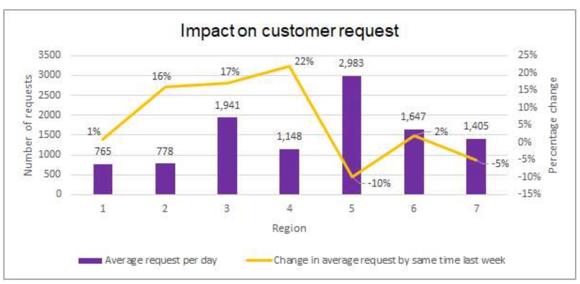


Figure 5: New pricing scheme yields changes in the average number of customer requests per day.

Firstly, the increase or decrease in average number of customer requests can prove our new pricing policy's advantage. According to data analysis, there is originally a total of over 10,600 customer requests in all seven regions. After the new model's application, region 1 and 6 witness minor increase while region 2, 3, and 4 experience a boost in the number of rider requests. The largest increment is in region 4 and hits 22%. There are minor deductions in the two other regions. However, the overall number still escalates by 379 requests, equivalent to 3.55% of the history statistic.

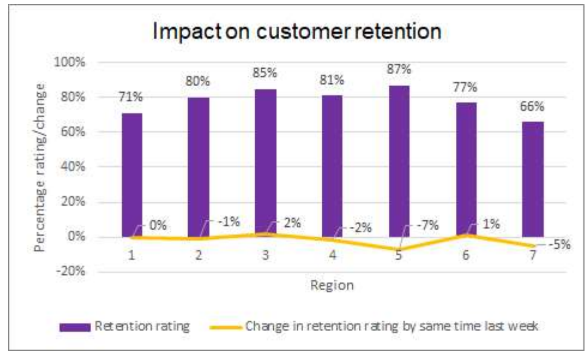


Figure 6: Customer retention is slightly adjusted under new pricing scheme.

Next, retention rate is the percentage of customers the company retains over the studied period. This indicator of customer loyalty has slightly dropped in many regions since the pricing adjustment strategy. In some regions, reduction in the number of customer requests results in customer retention rate. However, such declines are relatively small and negligible considering the maximum allowable percentage of decrease in customer retention equal 10%.

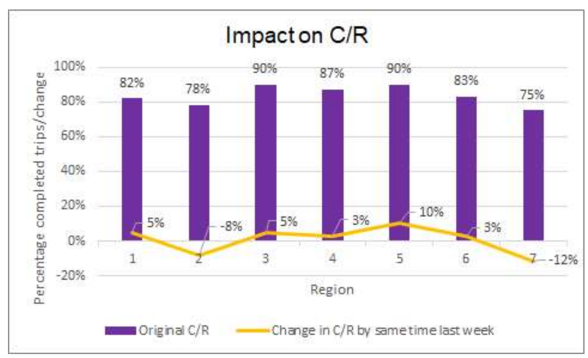


Figure 7: Impact of new pricing scheme on C/R ratio

Additionally, investigated under heavy traffic condition, the ratio of completed trips to total customer requests only lies at an average level of over 80%, reaching 90% in some regions. For region 7 whose decreased requests and decreased retention rate, its C/R ratio falls off by the largest amount and equal 12%. Meanwhile, region 4 whose highest demand growth witnesses rise in C/R ratio as well. This can be explained that when there is a surge in demand, customer biggest concern is no longer the service price. In fact, customers are willing to pay for excessively high price as long as their demand can be delivered. is no longer the key factor that matters.

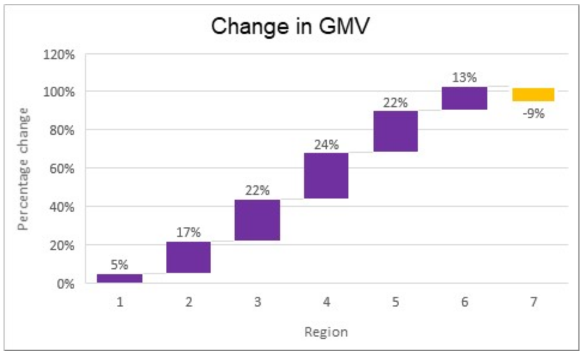


Figure 8: GMV grows considerably in most regions due to new pricing scheme.

As for GMV, GMV in six out of seven regions are appreciably improved. Especially in region 3, 4, and 5, it amounts to 22% and 24%. Region 7 is the only area that suffers from an abatement of 9% in GMV due to decreased requests and C/R ratio.

Based on the above analysis, we find that our proposed mathematical dynamic pricing model is eminently helpful for a company in optimizing revenue to over 20% during high demand period. Moreover, for an equal or less than 10% of customer retention rate, the model provides suitable price multiplier or a satisfactory service price.

5. Conclusion and Direction for Future Research

Given the world complicated economic situation in modern days, firms are strongly encouraged to have flexible pricing strategies that can adapt quickly to changes in marketplace using online environment. However, dynamic pricing decisions should be carefully made so as not to ruin customer experience. In this study, we present a new dynamic pricing strategy that acquires more input data of competitor's prices, customer expectations and traffic condition other than demand and supply only as in previous studies. Real-life surveys are executed before and after the practice of our price model to examine its impact on the business's profit as well as customer responses. The results show that under new price scheme, revenue and profit can be efficiently increased while customer satisfaction is virtually remained.

This study contributes to price optimization and create robust platforms for ride-hailing companies. Even though the business's GMV has siginificant growth, there is a relatively small decrease in customer retention rate. To grow a sustainable business, we recommend future studies to enhance our model by developing additional policies and strategies in order to keep customers coming back to the platform. In addition, increasing C/R ratio is another way to overcome limitations of our current model. Research on dispatching should be done to optimize this ratio of completed trips over total customer requests.

Furthermore, the mathematical model we introduce to compute dynamic price for ride-sharing platforms in this study is built upon many factors with alternate weightages that are researched to be appropriate at the present time. However, constants such as maximum acceptable price multiplier by customers – 3.7, maximum ratio of demand to supply nearby – 10, and maximum allowable customer retention rate – 10% should be regularly revised based on historical data as well as new market trends to avoid being outdated and to perfectly target company's orientations.

Reference

ABI Research. (2019). URL https://www.abiresearch.com/press/apac-ride-hailing-apps-go-on-steroids-to-offer-super-apps-in-effort-to-offset-universal-losses/

Bimpikis, K., Candogan, O., and Saban, D. (2019) Spatial pricing in ride-sharing networks. *Operations Research*.

Chen, L., Mislove, A., and Wilson, C. (2015) Peeking beneath the hood of uber. *Proceeding of the 2015 Internet Measurement Conference.*

Chen, M.K. and Sheldon, M. (2015) Dynamic pricing in a labor market: surge pricing and flexible work on the uber platform. *The 2016 ACM Conference.*

Google & Temasek/Bain, e Conomy SEA 2019. URL https://www.blog.google/documents/47/SEA_Internet_Economy_Report_2019.pdf/

Iglesias, R., Rossi, F., Zhang, R., and Pavone, M. (2019) A bcmp network approach to modeling and controlling autonomous mobility-on-demand systems. *The International Journal of Robotics Research*, 38(2-3):357–374.

Jin, Y., Ye, X., Liu, W., Wang, T., and Wang, H. (2019) Dynamic pricing model for cruising taxicab based on system dynamics. *Advances In Mechanical Engineering*.

Lei, C., Jiang, Z., and Ouyang, Y., (2019) Path-based dynamic pricing for vehicle allocation in ridesharing systems with fully compliant drivers. *Transportation Research Procedia*.

Lou, Q. and Saigal, R. (2017) Dynamic pricing for on-demand ride-sharing: a continuous approach. *SSRN Electronic Journal*.

Pandit, V.N, Mandar, D., Hanawai, M.K, and Moharir, S., (2019) Pricing in ride sharing platforms: static vs dynamic strategies. *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*.

Riquelme, C., Banerjee, S., and Johari, R. (2015) Pricing in ride-share platforms: a queuing-theoretic approach. *SSRN Electronic Journal.*

Yan, C., Zhu, H., Korolko, N., and Woodard, D. (2019) Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics, Forthcoming.*

Zhong, J., Lin, Y., and Yang, S. (2020) The impact of ride-hailing services on private car use in urban areas: an examination in chinese cities. *Journal of Advanced Transportation.*

USING K-NEAREST NEIGHBOR FOR FALL DETECTION A FEATURE SELECTION FRAMEWORK USING GREY WOLF OPTIMIZER

Phuong Thao Nguyen[1] and Vu Truong Son Dao[2]

[1] School of Industrial Engineering and Management, International University, Vietnam National University Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam, dvtson@hcmiu.edu.vn

[2] School of Industrial Engineering and Management, International University, Vietnam National University Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam, npthao130699@gmail.com

Corresponding author: Vu Truong Son Dao (dvtson@hcmiu.edu.vn)

Abstract. Human fall is a significant health accident which can cause serious injuries for many elderly people. This study focuses on developing a machine learning model which could detect the falling activities based on K-Nearest Neighbors classifier. However, a typical fall detection dataset contains many inputs, some of which are redundant and irrelevant; thus, affects the classification performance and processing speed. Feature selection is an important preprocessing step which is normally used to find the most important input features to reduce the classifier complexity and computational requirement. We use a recent swarm intelligence algorithm called Grey Wolf Optimizer (GWO), which mimicked the social hierarchy and hunting behavior of grey wolves, to search the available feasible space for this problem. Firstly, GWO was used to find the best features in the dataset. Secondly, the fitness function of GWO was evaluated by using K-Nearest Neighbor classifier (k-NN). We have obtained results with an accuracy of over 98 percent and an F1 score also over 0.98 points. Furthermore, the classifier requires much less time to train due to the reduced list of features.

Keywords: k-Nearest Neighbor, fall detection, feature selection, grey wolf optimization.

1. Introduction

Human Activity Recognition (HAR) is the process of supervising and analyzing the behavior of the user and the surrounding field to predict or detect basic daily activities, for instance, standing, lying, walking down/up, ascending/descending stairs, and most importantly, emergent events like falling (Reining et al. 2019). One of the important objectives of HAR is to provide all information related to human activity to assist users in their workplace and optimize the process. Falling also should be considered as one of the sudden activities in daily life when people perform other works.

Falling should be considered as one of the sudden activities in daily life (Chaudhuri, S., Thompson, H., & Demiris, G. 2014). The consequences of falls are unpredictable. Falls are not only a cause of death but also a factor that reduces the quality of life, Chaudhuri, S., Thompson, H., & Demiris, G. (2014), limits the ability to walk, and affects the quality of daily life. The paper indicated that falls are a very common unexpected accident in the elderly. Within a year, one in every 3 people over the age of 65 has a fall. Fall-related injury causes approximately 21,000 deaths in the United State.

This topic has already had much state-of-the-art research; therefore, it may be addressed in a variety of ways (Jobanputra, C., Bavishi, J., & Doshi, N. 2019). There are three main approaches, each with its own set of advantages and disadvantages, including the use of ambient/vision-based sensors to monitor human behavior, the use of wearable sensor devices to capture the signal of a human body gesture, and the use of a hybrid base-sensor. This paper is limited to the analysis of datasets derived from wearable sensor devices.

We will utilize a k-Nearest Neighbor (KNN) model in conjunction with preprocessing techniques to identify falling in this research. Additionally, a grey wolf optimization (GWO) method based on metaheuristics was used in the KNN model to enhance performance and decrease training time.

When developing a classification model, feature selection (FS) refers to the process of reducing the number of input variables in order to address the issue of high dimensionality. In FS, we choose a limited number of critical features and often discard irrelevant or noisy data to facilitate further analysis. Research about fall detection usually applied feature extraction instead of feature selection. For example, according to Chatzaki et al. (2017), the research "Human Daily

Activity and Fall Recognition Using a Smartphone's Acceleration Sensor", they extracted new features from the original dataset, separated them into 2 new datasets, and the maximum F1-score that they reached was just 0.835 over 1. Another research of Habib et al. (2014), which is "Smartphone-Based solutions for Fall Detection and Prevention: Challenges and Open Issues" used feature extraction to compute the threshold value such as mean, standard deviation, maximum, minimum value by threshold-based algorithms.

The Grey Wolf Optimization (GWO) is one of the fast-processing optimization algorithms based on the intelligence of swarms of animals that is attracting academics' attention for its ability to tackle a variety of hard practical issues (Mirjalili et al. 2014). In additionally, according to Cao et al. (2016), Swarm Intelligence (SI) techniques are increasingly being used to solve Feature Selection (FS), since SI is capable of addressing NP-hard computational issues. This algorithm is motivated by grey wolves' intellect and their unique ability to hunt in swarms.

2. Research Method

This system consists of two main steps. Firstly, data pre-processing, which included a statistical summary of the dataset, normalization, and Pearson correlation to determine the association between features and labels, as well as the relationship between features and labels. Secondly, grey wolf optimizer is used to optimize the architecture of KNN for classifying falling activity. This section describes the technique in detail.

2.1. Data collection

In this research, we used UP-Fall Detection Datasets from MDPI or Multidisciplinary Digital Publishing Institute which is a publisher of open access scientific journals (Table 2). The dataset was experienced and retrieved from 17 people that performed non-falling and fall activities. They used 5 wearable sensors to collect accelerometer, gyroscope, and ambient light data. Besides, they also applied ambient sensors and vision devices to acquire data.

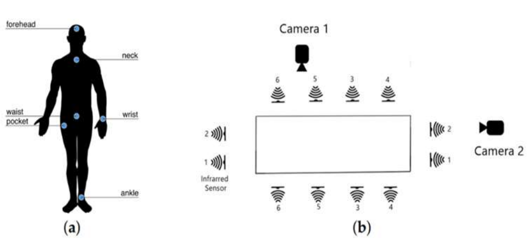


Figure 1. Distribution of the sensors.

(**a**) Wearable sensors and EEG headset located at the human body;
(**b**) Layout of the context-aware sensors and camera views.

Table 1. Detail of component sensors

Name	Quantity	Sensor
Raspberry	3	Infrarrojo
IMU	6	Accelerometer, Angular-Velocity, Luminosity
NeuroSky MindWave Mobile	1	Brainsensor
Infrarred Sensor (ABT 60)	6	Infrarred Sensor
LifeCam Cinema HD de Microsoft	2	Camera

However, in this paper, we focus on developing the model based on wearable-sensor devices. Hence, the data was discarded for the extraneous features including columns 2 to columns 36 and column 43. In addition, the data has also removed column 1 which indicates the timestamp of each experiment, because it didn't affect whether it is falling or non-falling activities. The last column named "isFall" which represented the label of each row. Specifically, if the column "isFall" is equal to 1 that means these features belong to fall activities, and vice versa.

2.2. Data preprocessing

2.2.1. Detecting missing value & standardise

In this section, the model was used Pandas info() function which display the total number of non-null observations present including the total number of entries. Once number of entries is not equal to number of non-null observations, we can begin to suspect missing values. In this dataset (figure 2), the number of entries is equal to the number of non-null observations, therefore, there is no null or missing values that need to be dropped or filled.

Table 2. List of original features in UP-Fall Dataset

Column	Features
[Column 1]	Timestamp
[Columns 2 - 4]	IMU (ankle): accelerometer X, Y & Z
[Columns 5 - 7]	IMU (ankle): angular velocity X, Y & Z
[Column 8]	IMU (ankle): luminosity
[Columns 9 - 11]	IMU (Right pocket): accelerometer X, Y & Z
[Columns 12 - 14]	IMU (Right pocket): angular velocity X, Y & Z
[Column 15]	IMU (Right pocket): luminosity
[Columns 16 - 18]	IMU (belt/waist): accelerometer X, Y & Z
[Columns 19 - 21]	IMU (belt/waist): angular velocity X, Y & Z
[Column 22]	IMU (belt/waist): luminosity
[Columns 23 - 25]	IMU (neck): accelerometer X, Y & Z
[Columns 26 - 28]	IMU (neck): angular velocity X, Y & Z
[Column 29]	IMU (neck): luminosity
[Columns 30 - 32]	IMU (wrist): accelerometer X, Y & Z
[Columns 33 - 35]	IMU (wrist): angular velocity X, Y & Z
[Column 36]	IMU (wrist): luminosity
[Column 37]	Raw value from EEG NeuroSky helmet
[Column 38]	Infrared Sensor 1
[Column 39]	Infrared Sensor 2
[Column 40]	Infrared Sensor 3
[Column 41]	Infrared Sensor 4
[Column 42]	Infrared Sensor 5
[Column 43]	Infrared Sensor 6
[Column 43]	isFall

2.2.2. Pearson correlation

In the Pearson correlation coefficient (r) was applied which is a common method and time. This coefficient is a test statistic that measures the relationship or association between features to each other and between features and labels (Correlation, Pearson. 2017).

Pearson correlation coefficient (r) fluctuates in the continuous range from -1 to +1:

r = 0: Two features have no linear correlation

r = 1; r = -1: Two features have an absolute linear relationship.

r < 0: Negative correlation coefficient. That is, the value of feature x increases, the value of feature y decreases and vice versa, the value of feature y increases, the value of feature x decreases.

r > 0: Positive correlation coefficient. That is, the value of feature x increases, the value of feature y increases and vice versa, the value of feature y increases, the value of feature x also increases.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 197645 entries, 0 to 197644
Data columns (total 36 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   x_ankle(a)          197645 non-null  float64
 1   y_ankle(a)          197645 non-null  float64
 2   z_ankle(a)          197645 non-null  float64
 3   x-ankle(deg/s)      197645 non-null  float64
 4   y-ankle(deg/s)      197645 non-null  float64
 5   z-ankle(deg/s)      197645 non-null  float64
 6   illuminance (lx)    197645 non-null  float64
 7   x-RP(a)             197645 non-null  float64
 8   y-RP(a)             197645 non-null  float64
 9   z-RP(a)             197645 non-null  float64
 10  x-RP(deg/s)         197645 non-null  float64
 11  y-RP(deg/s)         197645 non-null  float64
 12  z-RP(deg/s)         197645 non-null  float64
 13  illuminance (lx).1  197645 non-null  float64
 14  x-Belt(a)           197645 non-null  float64
 15  y-Belt(a)           197645 non-null  float64
 16  z-Belt(a)           197645 non-null  float64
 17  x-Belt(deg/s)       197645 non-null  float64
 18  y-Belt(deg/s)       197645 non-null  float64
 19  z-Belt(deg/s)       197645 non-null  float64
 20  illuminance (lx).2  197645 non-null  float64
 21  x-Neck(a)           197645 non-null  float64
 22  y-Neck(a)           197645 non-null  float64
 23  z-Neck(a)           197645 non-null  float64
 24  x-Neck(deg/s)       197645 non-null  float64
 25  y-Neck(deg/s)       197645 non-null  float64
 26  z-Neck(deg/s)       197645 non-null  float64
 27  illuminance (lx).3  197645 non-null  float64
 28  x-Wrist(a)          197645 non-null  float64
 29  y-Wrist(a)          197645 non-null  float64
 30  z-Wrist(a)          197645 non-null  float64
 31  x-Wrist(deg/s)      197645 non-null  float64
 32  y-Wrist(deg/s)      197645 non-null  float64
 33  z-Wrist(deg/s)      197645 non-null  float64
 34  illuminance (lx).4  197645 non-null  float64
 35  isFall              197645 non-null  int64
dtypes: float64(35), int64(1)
memory usage: 54.3 MB
```

Figure 2. Display of the total number of non-null observations and entries

Between ± 0.50 and ± 1: strong correlation

Between ± 0.30 and ± 0.49: Medium correlation

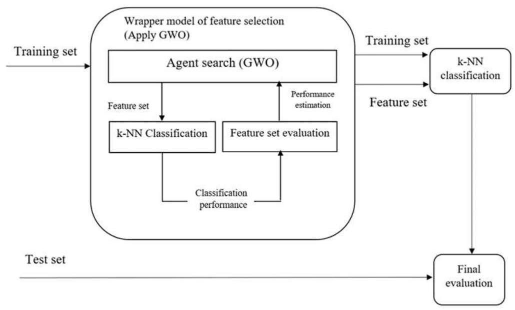
Below ± 0.29: Small correlation

We can see the figure 4 that most of the feature is independent with the others and does not have strong correlation with each other, or with the label, except for the correlation between "z-Belt(a)" and "z-RP(a)", and between "z-Belt(a)" and "z-Neck(a)"**.** The coefficient between "z-Belt(a)" and "z-RP(a)" is equal 0.59. The coefficient between "z-Belt(a)" and "z-Neck(a)" is equal 0.59. The feature "z-Belt(a)" mean the wearable sensor was worn on the belt. The feature "z-RP(a)" mean the wearable sensor was placed inside the right pocket of participants. The feature "z-Neck(a)" mean the wearable sensor was worn around the neck. This result shows that we can remove 1 of the 2 features that have a high correlation coefficient value. Because when one of the two features changes, the other feature will also change in the same direction.

2.3 Proposed framework

As demonstrated in Figure 3, GWO was applied in Feature selection. Unlike other traditional feature selection algorithms which they only focus on reducing the number of features as little as possible, the objectives of applying GWO includes 2 points:

- Find the smallest features subset.
- Optimize classification performance.



.

Figure 3. Flowchart of the proposed solution

2.3.1 Grey wolf optimizer (GWO)

GWO is the swarm-intelligent metaheuristic whose approach is to maintain and improve multiple candidate solutions, often using population characteristics to guide the search and find out the most optimal solution (Mirjalili et al. 2014). This optimization is applied to the study because of two reasons. Considering the fall detection problem, the goal is to minimize the number of features to reduce the time of the training model and maximize the classification performance.

In GWO, the best prey position as computed by alpha in cooperation with the beta, delta is considered as the best number of features subset for the feature selection. The GWO algorithm is inspired by the hunting techniques used by the Grey Wolve and the social hierarchy.

❖ About the rank of wolves in pack (Figure 4):

Alpha wolf: pack leader or in mathematical model means fittest solution.

Beta wolf: best candidate to be Alpha wolf or in mathematical model means second fitness solution.

Delta wolf: Provide food to the Pack & work for Pack in case of any danger or in mathematical model means third fitness solution

Omega: Omega is the Grey wolf's lowest rank. They are frail wolves who must depend on the pack's other wolves.

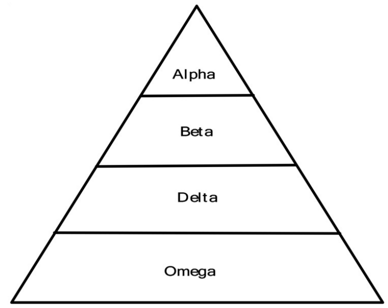


Figure 4. Social hierarchy of wolf in GWO

❖ Main steps of Grey Wolf Hunting are:

Searching: Searching for the Prey.

Pursuing: Tracking, chasing, and approaching the Prey.

Encircling and Harassing the Prey until it stops moving.

Attack: Attacking the Prey.

2.3.2 Pseudocode GWO

After pre-processing the data, we use Grey Wolf Optimize (GWO) to optimize the effectiveness of the k-Nearest Neighbor (KNN) for identifying whether the activity is falling. In this part, we utilized GWO to determine the optimum set of features that will classify by KNN classifier. The optimal answer is referred to as alpha when constructing a mathematical model. Beta and delta are regarded to be the second and third answers.

Begin

Initialize the paramters *objf, lb, ub, dim, popSize, Iter, trainInput, trainOutput*

where

‘’’Objf: objective function means fitness function

Lb: Lower bound (-1)

Ub: Upper bound (1)

Dim: numFeaturesData

popSize: size of population,

Iter: maximum number of iterations

trainInput

trainOutput’’’

Generate the initial positions of Grey Wolves with ub and lb;

Calculate the *fitness funtion* (Equation: 1) of initial position for each Grey Wolf;

Alpha_score: the grey wolf with the first maximum fitness;

Beta_score: the grey wolf with the second maximum fitness;

Delta_score: the grey wolf with the third maximum fitness;

For Iteration **in range** Max_iter

For i=0 **in range** popSize

Update *a* (Equation: 13)

Calculate the *fitness function* of all Grey Wolves with selected features;

Update *Alpha_score, Beta_score, Delta_score*;

Update $\vec{A}, \vec{C}, \vec{D}$ of each Grey Wolf (Equation: 2-9)

Update the position of current Grey Wolf (Equation: 12)

end for

Return the selected features of *Alpha_pos* as optimal number of features subset

2.3.3. Explanation of equation

Fitness Function

Because the objective includes in not only reducing the features but also improving classification efficiency. Therefore, this research decided to take the formular (Equation 1) for calculating the accuracy of the classification as the fitness function. The accuracy of the k-NN model that will show whether our model is efficiency, and the number of features after we want to reduce. I associate each factor a proportion as below.

$$99\% \,.\, (1 - \text{Accuracy}) + 1\% \,.\, \left(\frac{\sum Reduced\ Features}{35}\right) \quad (1)$$

Equation (2), (3) illustrate the GWO's phase of encircling prey:

$$\vec{D} = \left|\vec{C} \,.\, \overrightarrow{X_p}(t) - \vec{X}(t)\right| \quad (2)$$

$$\vec{X}(t+1) = \overrightarrow{X_p}(t) - \vec{A}(\vec{D}) \quad (3)$$

The coefficient vectors: $\vec{A}, \vec{C}, \vec{D}$:

$$\vec{A} = 2\vec{a} \,.\, \vec{r_1} - \vec{a} \quad (4)$$

$$\vec{C} = 2 \,.\, \vec{r_2} \quad (5)$$

$\vec{r_1}$, $\vec{r_2}$: random vectors in [0,1]

Equation (6-12) determine the wolf's last location (Hunting behavior):

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - A_1(\overrightarrow{D_\alpha}) \quad (6)$$

$$\overrightarrow{X_2} = \overrightarrow{X_\beta} - A_2(\overrightarrow{D_\beta}) \quad (7)$$

$$\overrightarrow{X_3} = \overrightarrow{X_\delta} - A_3(\overrightarrow{D_\delta}) \quad (8)$$

$\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$, $\overrightarrow{X_\delta}$: is the vector of the prey's position

$\overrightarrow{X_1}$, $\overrightarrow{X_2}$, $\overrightarrow{X_3}$: is the vector of the grey wolf's position

Where the distance between each wolf and its prey is described in equations (7,8,9):

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1} \,.\, \overrightarrow{X_\alpha}(t) - \bar{X}(t)\right| \quad (9)$$

$$\overrightarrow{D_\beta} = \left|\overrightarrow{C_2} \,.\, \overrightarrow{X_\beta}(t) - \bar{X}(t)\right| \quad (10)$$

$$\overrightarrow{D_\delta} = \left|\overrightarrow{C_3} \,.\, \overrightarrow{X_\delta}(t) - \bar{X}(t)\right| \quad (11)$$

Position of Grey Wolves:

$$\vec{X}(t+1) = \frac{\overrightarrow{X_1 + X_2 + X_3}}{3} \quad (12)$$

Linear function of parameter a

This purpose of this formula is that controlled trade-off between the exploration stage and exploitation stage. The vector a decrease over iterations linearly from 2 to 0. When the vector a decrease, it leads to the $\vec{A}$ also decreases.

$$\vec{a} = 2 - \text{I} \,.\, \left(\frac{2}{\text{Max\_iter}}\right) \quad (13)$$

Where,

I: current Iteration

Max_Iter: Maximum of Iteration that allowed

2.3.4. K-Nearest Neighbor (KNN)

The **x** represents the predictor, and **y** denote the class which include in falling or not-falling activity. The k-NN classifier predicts the class of **x** as the majority class among its **k** nearest neighbors (Guo et al. 2003). The majority class or similarity define in k-NN means that the distance between the new point data to test and those n_i points in training data set. In this model, we apply the Euclidean formular (Equation 14) to estimate this distance.

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + \ldots + (x_n - x'_n)^2} \quad (14)$$

For example, in figure 5, first we need to calculate the distance between new point to be tested and those n points in training data set which is 15 points in this example. Then, we will choose 3 nearest neighbors (k=3), or 5 nearest neightbors (k=5). After that we will choose the class contain higher density based on nearest neighbor density estimation algorithm.

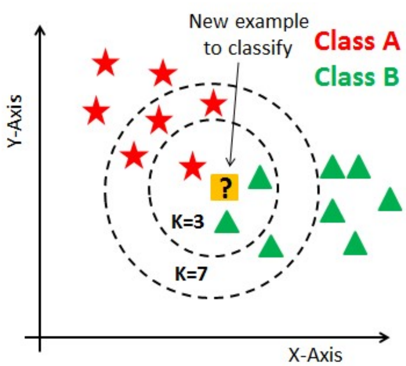


Figure 5. Example of k-NN algorithm

3. Results and analysis

3.1. Performance evaluation

3.1.1. K-Fold cross validation

The Cross validation is a sampling technique for evaluating machine learning models in cases where the data is not very abundant (Berrar, D. 2019). The key parameter in this technique is ***k***, which represents the number of groups into which the data will be split. For that reason, it is called k-fold cross-validation. When the value of k is chosen, that value is used directly in the name of the evaluation method. In this study, **k=10** will be applied.

3.1.2. F1-Score

Confusion matrix is a method of evaluating the results of classification problems with consideration of both the accuracy and generality of predictions for each class. A confusion matrix consists of the following four metrics for each classifier:

- True Positive (TP): the number of activities which are correctly presumed to fall.
- True Negative (TN): the number of activities which are correctly presumed to not-falling.
- False Positive (FP): the number of activities which are not correctly presumed to fall.
- False Negative (FN): the number of activities which are not correctly presumed to not-falling.

Table 3. Confusion matrix

Truth	Not-Falling (0)	**TN**	**FP**
	Falling (1)	**FN**	**TP**
		Not-Falling (0)	Falling (1)
		Predicted	

From these 4 indicators, we have 2 numbers to evaluate the reliability of a model: Precision and Recall. To evaluate the overall reliability of the model, the Precision and Recall indexes were combined into a single index: F-score, calculated by the calculated by the equation 4.

A model has a high F-score only when both Precision and Recall are high. Either one of these two metrics being low will drag down the F-score. The worst case when one of the two stats Precison and Recall is 0 will bring the F-score to 0. The best case when both stat points reach a value of 1, then the F-score will be 1.

x represents the predictor, and **y** denote the class which include in falling or not-falling activity. The k-NN classifier predicts the class of **x** as the majority class among its **k** nearest neighbors (Guo et al. 2003). The majority class or similarity define in k-NN means that the distance between the new point data to test and those n_i points in training data set. In this model, we apply the Euclidean formular (Equation 14) to estimate this distance.

3.2. Analysis results using various k value and test size (K-NN)

According to Table 4, the best results are obtained when k equals 1 and the test size is 0.2. The confusion matrix of the best result illustrates in the figure 6, The test size is 0.2, which indicates that 80 percent of the original dataset will be split into a training dataset and the remaining 20 percent will be used to evaluate the effectiveness of a machine learning algorithm when it is used to examniate predictions on dataset that was not used to train the model. Additionally, we can observe that setting k to an odd constant often produces a greater result than setting it to an even constant. It is also understandable that when k is odd, the k-NN method will easily find the bulk of the target group among the nearest neighbors. Therefore, the k=1, and test size is equal 0.2 would be used when running GWO.

Table 4. The summary of results when applying various k values and test size.

k	Test size	Accuracy 10 times from k-Fold	Standard Deviation k-Fold	True Positive	True Negative	False Positive	False Negative	Precision	Recall	F1 Score
1	0.2	99.35	0.06	6275	33011	60	183	0.991	0.972	0.981
2	0.2	99.01	0.08	6107	33036	35	351	0.994	0.946	0.969
3	**0.2**	**98.99**	**0.08**	**6188**	**32960**	**111**	**270**	**0.982**	**0.958**	**0.97**
4	0.2	98.77	0.08	6057	33005	66	401	0.989	0.938	0.963
1	**0.25**	**99.32**	**0.08**	**7824**	**41267**	**93**	**228**	**0.988**	**0.972**	**0.98**
2	0.25	98.95	0.1	7623	41307	53	429	0.993	0.947	0.969
3	**0.25**	**98.93**	**0.09**	**7728**	**41205**	**155**	**324**	**0.98**	**0.96**	**0.97**
4	0.25	98.7	0.08	7560	41274	86	492	0.989	0.939	0.963
1	**0.3**	**99.28**	**0.08**	**9380**	**49486**	**136**	**292**	**0.986**	**0.97**	**0.978**
2	0.3	98.91	0.1	9135	49551	71	537	0.992	0.944	0.968
3	**0.3**	**98.88**	**0.09**	**9260**	**49417**	**205**	**412**	**0.978**	**0.957**	**0.968**
4	0.3	98.62	0.11	9054	49499	123	618	0.987	0.936	0.961

3.1. Experimental results of GWO-KNN

The model was taken the experiences with different values of population size and iteration size (Table 3.3). The result of dataset with 16 features are slightly better than those with 20 or 18 features. Moreover, regarding to the table 5, which indicate when the population size rises, the ratio of properly predicted positive points to total predicted points (F1-Score) decreased. Additionally, when the number of Iteration sizes increases, the predictive effectiveness of the model improves. Typically, the same population size is 20, and when the number of iterations is raised from 500 to 1500, even if the number of features remains equal to 16, the model's performance assessment increased

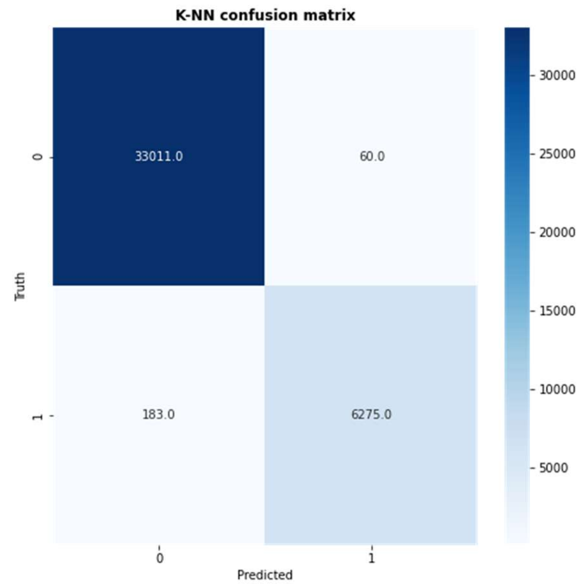


Figure 6. The confusion matrix of the result (k=1, test size=0.2)

by 13 percent and the AUC score increased by 8.23 percent. Alternatively, when the population size remains constant at 40 but the number of iterations is increased from 500 to 1000, the F1-score rises by 8,159 percent and the AUC score also increased by 4.9 percent.

Table 5. The summary of evaluating performance when applying various population size and iteration size

k	Population size	Iteration size	Optimum number of features	F1 Score	AUC	Cohens Kappa
1	20	3	16	0.96395	0.9739	0.95701
1	20	500	16	0.95608	0.96898	0.94476
1	**20**	**1500**	**16**	**0.971379**	**0.97908**	**0.96586**
1	40	500	20	0.820749	0.884941	0.786895
1	40	1000	18	0.960518	0.96091	0.95299

The study presents research observations obtained using Grey Wolf Optimization (GWO) on the k-Nearest Neighbor classifier (GWO – KNN). The table 6 illustrates 16 chosen features from the dataset created using GWO with the best fitness is at 0.0098714264. Moreover, the number of chosen characteristics obtained the greatest classification accuracy when compared to all other factors. This method requires less training time than the others.

Table 6. The list of features in new datasets when applying GWO

x_ankle(a)	Selected
y_ankle(a)	0
z_ankle(a)	0
x-ankle(deg/s)	0
y-ankle(deg/s)	Selected
z-ankle(deg/s)	0
illuminance (lx)	Selected
x-RP(a)	0
y-RP(a)	Selected
z-RP(a)	Selected
x-RP(deg/s)	0
y-RP(deg/s)	Selected
z-RP(deg/s)	Selected
illuminance (lx)	Selected
x-Belt(a)	0
y-Belt(a)	0
z-Belt(a)	0
x-Belt(deg/s)	0
y-Belt(deg/s)	0
z-Belt(deg/s)	0
illuminance (lx)	Selected
x-Neck(a)	0
y-Neck(a)	Selected
z-Neck(a)	Selected
x-Neck(deg/s)	0
y-Neck(deg/s)	0
z-Neck(deg/s)	Selected
illuminance (lx)	Selected
x-Wrist(a)	0
y-Wrist(a)	Selected
z-Wrist(a)	0
x-Wrist(deg/s)	0
y-Wrist(deg/s)	Selected
z-Wrist(deg/s)	Selected
illuminance (lx)	0

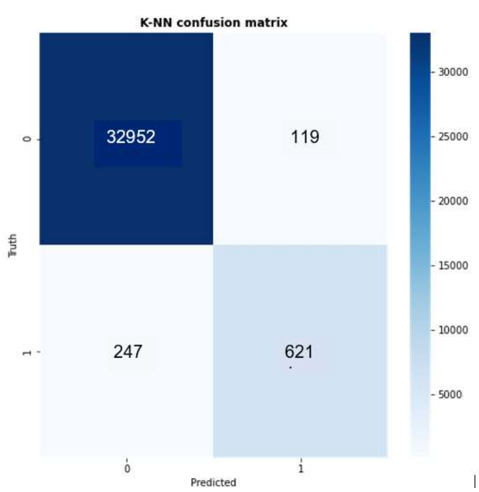


Figure 7. The confusion matrix of result (population size=20, iteration size=1500)

4. Conclusion

In this thesis, the GWO method was used to decrease the classifier complexity and processing requirements of a machine learning model capable of detecting falling. It has been shown that the suggested approach improves the model's accuracy and generates a new dataset with smaller data size. The system was evaluated using data from MDPI's UP-Fall Detection Datasets, a publisher of open access scientific publications. The system was evaluated using data

from MDPI's UP-Fall Detection Datasets, a publisher of open access scientific publications.

Finally, the performance of the GWO algorithm is outperformance than the traditional PCA and LDA algorithms. An overall comparison revealed that the GWO algorithm produces the most acceptable results with an average improvement of about 20% compare with the results of PCA, LDA algorithm. This may be seen as an encouraging sign to trial more of these ensemble techniques will be used in the future, with the goal of further improving the capacity to anticipate falling and human recognition activity.

In the future, this work can be developed to create practical applications for people or apply in the medical industry. Additionally, this approach may be applied to real-time data as it becomes available.

Reference

Abbate, S., Avvenuti, M., Corsini, P., Light, J., & Vecchio, A. (2010). Monitoring of Human Movements for Fall Detection and Activities Recognition in Elderly Care Using Wireless Sensor Network: a Survey. *Wireless Sensor Networks: Application-Centric Design*. Published. https://doi.org/10.5772/13802

Al-Tashi, Q., Rais, H., & Jadid, S. (2018). Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification. *Advances in Intelligent Systems and Computing*, 257–266. https://doi.org/10.1007/978-3-319-99007-1_25

Berrar, D. (2019). Cross-Validation. Encyclopedia of Bioinformatics and Computational Biology, 542–545. https://doi.org/10.1016/b978-0-12-809633-8.20349-x

Brezočnik, L., Fister, I., & Podgorelec, V. (2018). Swarm Intelligence Algorithms for Feature Selection: A Review. *Applied Sciences*, *8*(9), 1521. https://doi.org/10.3390/app8091521

Cao, J., Cui, H., Shi, H., & Jiao, L. (2016). Big Data: A Parallel Particle Swarm Optimization-Back-Propagation Neural Network Algorithm Based on MapReduce. *PLOS ONE*, *11*(6), e0157551. https://doi.org/10.1371/journal.pone.0157551

Correlation, Pearson. (2017). *The SAGE Encyclopedia of Communication Research Methods*. Published. https://doi.org/10.4135/9781483381411.n99

Chatzaki, C., Pediaditis, M., Vavoulas, G., & Tsiknakis, M. (2017). Human Daily Activity and Fall Recognition Using a Smartphone's Acceleration Sensor. *Communications in Computer and Information Science*, 100–118. https://doi.org/10.1007/978-3-319-62704-5_7

Chaudhuri, S., Thompson, H., & Demiris, G. (2014). Fall Detection Devices and Their Use With Older Adults. *Journal of Geriatric Physical Therapy*, *37*(4), 178–196. https://doi.org/10.1519/jpt.0b013e3182abe779

Cheng, S., Shi, Y., Qin, Q., & Bai, R. (2013). Swarm Intelligence in Big Data Analytics. *Intelligent Data Engineering and Automated Learning – IDEAL 2013*, 417–426. https://doi.org/10.1007/978-3-642-41278-3_51

Choi, Y., Ralhan, A. S., & Ko, S. (2011). A Study on Machine Learning Algorithms for Fall Detection and Movement Classification. *2011 International Conference on Information Science and Applications*. Published. https://doi.org/10.1109/icisa.2011.5772404

Faris, H., Aljarah, I., Mirjalili, S., Castillo, P. A., & Merelo, J. J. (2016). EvoloPy: An Open-source Nature-inspired Optimization Framework in Python. *Proceedings of the 8th International Joint Conference on Computational Intelligence*. Published. https://doi.org/10.5220/0006048201710177

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, 986–996. https://doi.org/10.1007/978-3-540-39964-3_62

Habib, M., Mohktar, M., Kamaruzzaman, S., Lim, K., Pin, T., & Ibrahim, F. (2014). Smartphone-Based Solutions for Fall Detection and Prevention: Challenges and Open Issues. *Sensors*, *14*(4), 7181–7208. https://doi.org/10.3390/s140407181

Jobanputra, C., Bavishi, J., & Doshi, N. (2019). Human Activity Recognition: A Survey. *Procedia Computer Science*, *155*, 698–703. https://doi.org/10.1016/j.procs.2019.08.100

Khan, A. M., Lee, Y. K., Lee, S. Y., & Kim, T. S. (2010). Human Activity Recognition via an Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis. *2010 5th International Conference on Future Information Technology*. Published. https://doi.org/10.1109/futuretech.2010.5482729

Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., & Peñafort-Asturiano, C. (2019). UP-Fall Detection Dataset: A Multimodal Approach. *Sensors*, *19*(9), 1988. https://doi.org/10.3390/s19091988

Mavrovouniotis, M., Li, C., & Yang, S. (2017). A survey of swarm intelligence for dynamic optimization: Algorithms and applications. *Swarm and Evolutionary Computation*, *33*, 1–17. https://doi.org/10.1016/j.swevo.2016.12.005

Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and approaches. *Neurocomputing*, *100*, 144–152. https://doi.org/10.1016/j.neucom.2011.09.037

Negi, G., Kumar, A., Pant, S., & Ram, M. (2020). GWO: a review and applications. *International Journal of System Assurance Engineering and Management*, *12*(1), 1–8. https://doi.org/10.1007/s13198-020-00995-8

Wu, F., Zhao, H., Zhao, Y., & Zhong, H. (2015). Development of a Wearable-Sensor-Based Fall Detection System. *International Journal of Telemedicine and Applications*, *2015*, 1–11. https://doi.org/10.1155/2015/576364

APPLICATION OF MACHINE LEARNING IN CREDIT CARD DEFAULT PREDICTION

Nga Vu[1], Son Dao[1]

[1] *School of Industrial Engineering and Management, VNU – International University, Ho Chi Minh city, Vietnam*

vvhangnga.1904@gmail.com, dvtson@hcmiu.edu.vn

Abstract. Risk prediction is an essential solution for banks to maintain a stable financial operation system and reduce uncertain consequences. Credit card default leads banks to suffer risks and losses whenever customers lately pay the limit balance debt. Hence, cases of default must be correctly classified in advance. This work proposes a machine learning model of Extreme Gradient Boosting (XGBoost), together with a feature selection of Grey Wolf Optimization (GWO) to predict credit card default cases. This paper also proposes hyperparameters tuning method to optimize prediction performance. Moreover, results achieved are compared with several machine learning algorithms such as Logistics Regression (LR), Decision Tree (DT) and K-Nearest Neighbor (KNN). Our proposed method not only reduces the number of features to only 15, but also achieves a higher predictive performance of 85.58% accuracy and 0.6736 in AUC score. This work has the possibility to be applied and becomes supportive for bank system.

Keywords: Credit Card, XGBoost, Hyperparameter, Feature Selection, Grey Wolf Optimization

1. Introduction

In bank system, Artificial Intelligence (AI) can be used to verify customers' identification or authentication, smoothen the banking process through voice detection, analyze insights or risk assessment. Machine learning, as an AI technique, if well implemented, will have a positive effect on enterprises' profit. Normally, banks have provided many services, such as international money transfer, receiving money transferred from abroad, insurance premium collection or property transfer payment services. However, although there are many opportunities for banks to be developed, there are still many risks expose in customers lending. For instance, from 2003 to 2007, the average charge-off rate in customer credit from all US lending institutions was just 4.72%. Though, in the third quarter of 2019, this rate increased to 10.1%. * Hence, there are also some aspects that machine learning has not been thoroughly explored to achieve the highest potential in banking, such as credit card defaults prediction.

In this paper, XGBoost approach is implemented, with a metaheuristics-based feature selection of Grey Wolf Optimization algorithm to improve the model performance. This approach will be benchmarked with other machine learning models, such as Logistics Regression (LR), K-Nearest Neighbor (KNN) and Decision Tree (DT) to evaluate its performance and efficiency. The proposed method will then be investigated to see whether it is applicable than other methods or not. Using a dataset from a bank in Taiwan from April to September 2005, we establish a model to predict and classify whether customers will lately pay for their credit card loan or not. The structure of this paper is as follows. Section 2 shows related works and literature of this research. In Section 3, methodology is showed, which describes the machine learning framework and its corresponding feature selection techniques. Section 4 presents the empirical results.

2. Literature Review

2.1. Credit Risk Prediction techniques

In our case study, credit risk prediction is to evaluate if a customer can fall into default status in the future or not, so they can continue to grant the loan. The result of this is the prediction of default rate of customers. Consequently, it is important for banks to have information about their customers and then can predict and assess their default. Hence, this helps banks to make decision whether they should continue to offer a loan to the customer or not. In some scenarios, big and complex data should be collected and proceeded with data mining techniques to gain a suitable outcome. The use of methods such as statistical regression, Decision Trees (CART), Neural Networks, and KNN have been reviewed in the paper

** Data available from the Federal Reserve Board at http://www.federalreserve.gov/releases/chargeoff/*

of Galindo and Tamayo (2000). Twala (2010) presented credit risk prediction at various attribute noise levels using five classifiers, including Artificial Neural Networks (ANN), DT, Naïve Bayes (NB), KNN, and Logistics discrimination (LD). The author tried to prove that the method of ensemble classifiers was a potential factor in improving the accuracy rate. Kruppa et al. (2013) also proposed another credit risk prediction focusing on Random Forests (RF), KNN, and bagged K-Nearest Neighbors (bNN) since these are fast and common to use. The authors also examined and proved these algorithms' effectiveness in many other problems.

2.2. Extreme Gradient Boosting (XGBoost)

Ye et al. (2017) said that boosting method is an effectively integrated learning algorithm. This can make a weak classification become stronger in order to gain a higher accurate classification outcome. Gradient Boosting (GB) is also boosting method, and this algorithm's goal is to make the loss function reduce when passing continuous iterations, and then can improve the robustness. Therefore, Extreme Gradient Boosting (XGBoost) depends on a decision tree, and it is a faster and upper level of GB. XGBoost is an upgraded improvement of the GB algorithm which produces a more accurate outcome. It uses both algorithms of linear and decision tree to build a model with higher accuracy. Moreover, XGBoost also can handle missing data in the dataset and avoid over-fitting. This new approach finds a new classifier that can reduce the loss function (Parsa et al, 2020).

Guo et al. (2019) implemented the XGBoost algorithm for fitness evaluation used in wearable running monitoring and uses the advanced feature selection and hyperparameter optimization. And the XGBoost algorithm outperforms others algorithm, having the accuracy score of 96.41% for males and 98.11% for females. Ryu et al. (2020) used XGBoost with derived feature extraction together with hyperparameter optimization to make a prediction model of dementia risk and had the accuracy of 85.61%, and F1-score of 79.28%. Li and Zhang (2019) proposed XGBoost classifier for predictive model in the case of orthopaedic auxiliary classification, which supports the doctor diagnosis decision, had the result of 95.1% accuracy and 92.8% recall rate and outperforms other classification models. Li et al. (2020) implemented the XGBoost model to personal loan problem, using the

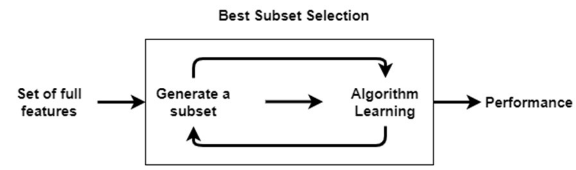


Figure. 1. Wrapper Feature Selection Method

Lending Club Platform dataset in USA. They showed that XGBoost gained advantages in both feature selection and classification performance, resulting in 93.7% accuracy and 94.81% AUC score compared to other models.

2.3. Feature Selection techniques

Yu et al. (2004) have separated features into four main types: irrelevant features, weakly relevant and redundant features, weakly relevant but non-redundant features, and strongly relevant features. The number of features in the dataset is in proportion to hypothesis space size, which means if the number of feature increase, hypothesis space size also increases. Set of features plays an essential role in deciding which hypothesis is suitable with the predictive models. The main goal of feature selection is figure out which is relevant features, and which are the irrelevant features. Then, eliminate irrelevant and redundant attributes to achieve good classification results.

There are three main kinds of feature selection techniques chosen depending on the interaction with the learning model. The three techniques are Filter, Wrapper, and Embedded Methods. El & Benhlima (2016) had stated that Filter methods do not involve learning algorithms and do not evaluate a feature because after ranking features, they will reject features having scores below a predetermined threshold. However, Wrapper, which involves learning algorithms that evaluate the accuracy performance and hence, lead to better solutions. Some of the wrappers are sequential feature selection algorithms (Jain and Zonger, 1997), recursive feature elimination (Yan and Zhang, 2015), and particle swarm optimization algorithms (Zhang et al., 2014). Wrapper feature selection techniques are clearly described in **Fig. 1**.

Zorarpacı and Ozel (2016) also stated that Wrapper provides better performance of result, although many feature selections based on metaheuristics are also proposed. Mirjalili et al. (2014) suggested Grey Wolf Optimizer (GWO) algorithm for any feature selection problems because GWO follows the three best solutions of the wolves while other algorithms mainly focus on only the best one. Emary et al. (2015) had compared GWO with PSO and GA to see their differences and effectiveness in selecting features. On the other hand, Too et al. (2018) came up with a suggestion of competitive binary GWO (CBGWO) in the case of classifying EMG signal.

3. Methodology

This section will explain the process of our framework. Data firstly is preprocessed by standardization and outlier removal. Then, XGBoost is applied together with the wrapper feature selection technique for model training. Moreover, hyperparameter tuning is applied to see whether this technique is an improvement or not. Finally, the proposed framework is taken into comparison with other algorithms to evaluate the performance.

3.1. Data Preprocessing

The dataset of credit card customer default, which is from Taiwan bank, is used in this paper. The dataset is token in UCI repository. There are 30,000 datapoints and 24 features, with 23 features are explanatory and 1 is used as label. The dataset description is as follow:

- **X1**: The limit balance *(NT dollar)*: includes the customer's credit and their family credit.
- **X2**: Sex *(1 is men; 2 is women).*
- **X3**: Education *(1 is graduated; 2 is under-graduated university; 3 is high school; 4 is others).*
- **X4**: Marital status *(1 is married; 2 is single; 3 is others).*
- **X5**: Age *(year).*
- **X6 – X11**: History of past payment. With X6 is the repayment status in September, and X11 is in April. *(-1 is pay duly, 1 is payment delay for one month; 2 is for two months, and so on, to 9 is for nine months and above).*
- **X12 – X17**: Amount of bill statement *(NT dollar).* With X12 is amount of bill statement in September. Continually, X17 is amount in April.
- **X18 – X23**: Amount of previous payment *(NT dollar)*. With X18 is amount paid in September and continually to X23 is amount paid in April.
- **X24**: default status of next month, which is binary data *(1 is default; 0 is non-default)*

Then, the standardization process is conducted to make the same format for the dataset. IQR (Interquartile Range) is also used to exclude the outliers of data.

3.1.1. Standardization

Different columns of data have different ranges of values, some will have negative values while some

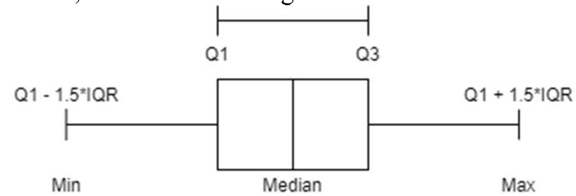


Figure. 2. Box-Whisker Plot showing IQR

will have positive ones; some will be integer whereas some will be decimal. Hence, if the data are not converted into one standard format, it is difficult for them to be compared and analyzed. Data standardization is the process of making data to have one common format and helping users to analyse and use it. The result of standardization is that all features will be rescaled to help the next outliers removing steps. The function of a Robust Standardization is shown below:

$$x_{stand} = \frac{x - mean(x)}{x_{75} - x_{25}} \quad (1)$$

Where $\boldsymbol{x}$ is the original feature vector. This technique is to rescale features value but still remain the value of label.

3.1.2. Interquartile Range (IQR)

To deal with a dataset, IQR is defined as a range between 25% and 75% sample quantiles Q_1 and Q_3, which is lower and upper quartile value, respectively. This covers the central 50% of data and the box-Whisker plot is showed below in **Fig. 2.**

The region between Q_1 and Q_3 is shown in a shape of box; two lines of whiskers are drawn. The IQR is the difference between Q1 and Q3:

$$IQR = Q_3 - Q_1 \quad (2)$$

The lower whisker limit is computed as:

$$Q_1 - 1.5 * IQR \quad (3)$$

The upper whisker limit is computed as:

$$Q_3 + 1.5 * IQR \quad (4)$$

3.2. Proposed Framework

3.2.1. Extreme Gradient Boosting

One of our research targets is to figure out the best combinations of features, which have the most performance in this binary classification. Assembly algorithms will help to produce and to assemble several classifiers which is individually weak but useful, to create a robust estimator. There are two ways to make this combination: bagging (random forests) and boosting. Natekin and Knoll (2013) stated that gradient boosting is sequentially constructed, which has a correlation with negative gradient of the loss function.

According to Chen and Guestrin (2016), XGBoost, which is an improvement from GBDT, haves two key optimization factors. The first factor is that it adds up a regularization term to objective function and then helps the model not fall into overfitting problems. While GBDT only use the first derivative, the second factor of XGBoost shows a second derivative Taylor expansion in optimization to the objective function. Hence, XGBoost defines the loss function more accurately. The loss function in the XGBoost, whose objective is to be minimized, is defined as follows:

$$L^{(t)} = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_t) \qquad (4)$$

Where: $\hat{y}_i^{(t)}$ is the prediction at t^{th} round while f_t is the structure of the tree. And n is number of observations. The regularization item is:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \qquad (5)$$

Where γ is the penalty coefficient, and the second part of this equation is the L2 norm of leaf node score. At t iterations, the model function is the $(t-1)^{th}$ prediction function added by a new tree. And the objective function is updated as by:

$$L^{(t)} = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(t)} + f_t(x_i)\right) + \sum_{i=1}^{t} \Omega(f_t) \qquad (6)$$

After applying Taylor second expansion, the objective function is as follows:

$$L^{(t)} = \sum_{i=1}^{n} \left[L\left(y_i, \hat{y}_i^{(t)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) \qquad (7)$$

Where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ is the first and second order derivative of the loss function.

The fitness function helps finding the maximum accuracy rate, which also help to evaluate the selected features. Fitness function is denoted as:

$$Fitness = \alpha P + \beta \frac{N-L}{L} \qquad (8)$$

Where P is the accuracy rate, L is the number of features selected, N is the total number of features. α and β are parameters that correspond with accuracy

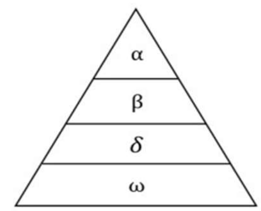


Figure. 3. Hierarchy of wolves in GWO

weight and quality of feature selection, $\alpha \in [0, 1]$ and $\beta = 1 - \alpha$.

3.2.2. Grey Wolf Optimization (GWO)

Grey Wolf Optimization (GWO) algorithm bases basically on the natural behavior of wolves in their wildlife habitat, especially their ways of hunting the prey. The organization system and the behavior of wolves are clearly described in (Mirjalili et al, 2014). The highest level of wolves is called *alpha*, whose responsibility is to manage and lead the pack in hunting, moving, sleeping time, and so on. Alpha may be not the strongest in physically health but be the strongest in intellectual problem solving. The second-highest level is *beta*, who is the direct subordinate of alpha. The lowest level in the hierarchy system is *omega* wolves, who have no subordinates. The other type of wolves, who is not alpha, beta, or omega, will be *delta* . The figure **Fig. 3** shows the clear hierarchy of the wolves.

As describe above, wolves encircle their prey in hunting. The position of wolves is updated by calculating the distance from the prey's updated position (estimated by α, β and δ) as follows.

$$\vec{D} = \left|\vec{C}.\vec{X}_p(t) - \vec{X}(t)\right| \qquad (9)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}.\vec{D} \qquad (10)$$

where t is the current iteration, $\vec{X}_p$ is the position vector of prey and $\vec{X}$ is the position vector of wolf. The coefficient vector $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = 2\vec{a}.\vec{r_1} - \vec{a} \quad (11)$$

$$\vec{C} = 2.\vec{r_2} \quad (12)$$

where $\vec{r}_1$ and $\vec{r}_2$ are random vector from 0 to 1. Parameter ***a*** is linearly decreasing (from 2 to 0), which is measured as below:

$$a = 2 - 2\left(\frac{t}{T}\right) \quad (13)$$

where t is the quantity of iterations, and T is the maximum quantity of iterations.

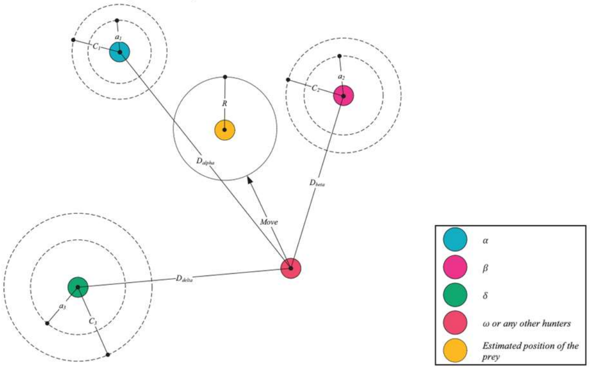


Figure. 4. Updating position in GWO (Mirjalili et al, 2014)

As can be seen in **Fig. 4**, grey wolf position is updated based on the prey position. Different positions surrounding the best agent will be gained based on the current position by modifying the value of $\vec{A}$ and $\vec{C}$ vectors. Hence, to describe the hunting behavior of grey wolves in mathematical way, the three best solutions are stored and applied to find other search agents.

Hence, it is important to attach the wolf position to a binary value and here we use BGWO method 1 (BGWO1) (Emary et al. 2016). This model uses a crossover method wolves update the position as below:

$$\Upsilon_i^{k+1} = Crossover(\gamma_1, \gamma_3, \gamma_3) \quad (14)$$

Where γ_1, γ_2, γ_3 all are binary values, which affected by the motion of three best positions α, β and δ, respectively. The binary vectors γ_1, γ_2, γ_3 are updated as followings:

$$\gamma_1^D = \begin{cases} 1, & (\Upsilon_\alpha^D + q_\alpha^D) \geq 1 \\ 0, & otherwise \end{cases} \quad (15)$$

Where in D dimension, γ_α^D is the position of alpha, and q_α^D is the binary step measured by:

$$q_\alpha^D = \begin{cases} 1, & p_\alpha^D \geq r \\ 0, & otherwise \end{cases} \quad (16)$$

Where, r is random number having value of 0 or 1, which is uniformly distributed. p_α^D represents the updating step in D dimension calculated by sigmoid transformation as below:

$$p_\alpha^D = \frac{1}{1 + \exp\left(-10 * (A_1^D D_\alpha^D - 0.5\right)} \quad (17)$$

Where, A_1^D and D_α^D is in continuous value range of grey wolf optimizer, which is already estimated in the above two equations.

```
Begin
Randomly initialize the population of grey wolves, X
Initialize the parameter, a, A⃗, C⃗
Evaluate fitness of wolves, F(X)
Set Xα, Xβ, Xδ as the position of highest, second and
third wolf.
for t = 1 to maximum number of iterations, T
    for i = 1 to number of wolf, N
        Compute γ1, γ2, γ3
        Generate Xi^new, by applying crossover between
        γ1, γ2, γ3.
next i
Evaluate the fitness of all grey wolves, F(Xi^new)
Update the position of α, β and δ
Update the parameter a, A⃗, C⃗
next t
```

Figure. 5. Pseudo-code for BGWO1

Similar with *alpha* equations given in the above function from (15) to (17), equations for *beta* and *delta* are similar. After having the result of γ_1^D, γ_2^D, γ_3^D the wolf position updating process through the crossover is given by:

$$\gamma^D = f(x) = \begin{cases} \gamma_1^D, & r < \frac{1}{3} \\ \gamma_2^D, & \frac{1}{3} \leq r \leq \frac{2}{3} \\ \gamma_3^D, & otherwise \end{cases} \quad (18)$$

where, γ_1^D, γ_2^D, γ_3^D are the first, second and third best binary values, γ^D is the output of dimension D, r denotes a random binary, which distributes uniformly.

At first, the position of grey wolves, X is randomly 1 or 0. After that, the fitness of each wolf is calculated. For each level of wolf, γ_1, γ_2, γ_3 are computed. Next, by using the crossover strategy, new position of the wolves is updated. The fitness of each wolf is then evaluated and the positions of best three ones α, β and δ are updated. This model will repeat continuously until certain iterations end. Lastly, alpha position is chosen to be best subset of features. According to Too et al. (2018), pseudo-code for BGWO1 is in **Fig. 5.**

3.2.3. Hyperparameter Tuning

Every ML algorithm have their own parameters based on the characteristic of that algorithm. In order to be more efficient, significant parameters need to be set with a certain value. A hyperparameter is not externally related to the process of training data and will not be changed during training time and the characteristics of data cannot measured its value.

Grid Search technique is applied in our case. 'n_estimators' is the number of boosted trees to fit,

Table 1. Performance of benchmark models

Classifier	LR	KNN	DT	XGBoost (optimized)
Accuracy	0.8477	0.8225	0.7406	0.8558
Precision	0	0.2268	0.2072	0.6666
Recall	0	0.0685	0.2487	0.0053
F1	0	0.1053	0.2261	0.0106

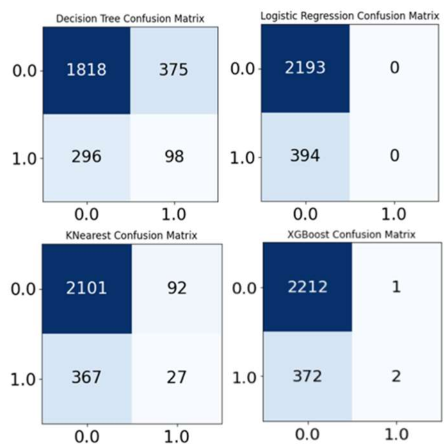


Figure. 6. Confusion matrix of benchmark models

while 'max_depth' represents the maximum number of splits; the higher value of this, the higher the chance model will overfit. 'min_child_weight' is used to control the over-fitting problem. The part of observations "subsample" can also prevent overfitting. In each iteration, 'learning rate' is used to modify tree impact and help the algorithm more robust (14). 'colsample_bytree' also help the model to prevent over-fitting. Moreover, after one boosting iteration, one base learner will be added and model is tuned through k-fold Cross-Validation, where k is set to equal to 5 and scoring is 'roc_auc'.

4. Results

The dataset is previously divided into with a ratio of 80:20, respectively. The evaluation metrics, which are given below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

$$Precision = \frac{TP}{TP + FP} \quad (20)$$

$$Recall = \frac{TP}{TP + FN} \quad (21)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (22)$$

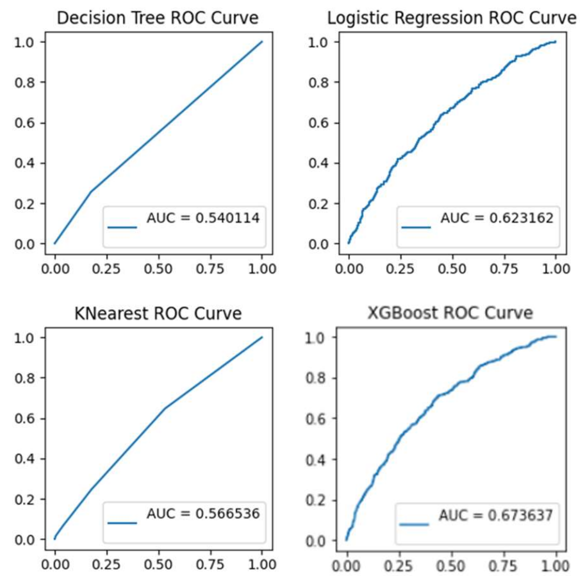


Figure. 7. Receiver Operating Characteristic (ROC) of benchmark models

Where: True Positive (TP) and True Negative (TN) are quantity of positive and negative instances that are correctly classified, respectively. On the other hand, False Positive (FP) and False Negative (FN) are quantity of positive and negative instances that are wrongly classified (Hossin and Sulaiman, 2015).

4.1. Analysis Results of Machine Learning Models

In this research, three classification algorithms such as LR, KNN, DT are firstly implemented. Before training these data, the removal of the outlier is applied. **Table 1**. Shows that LR classifier has the highest accuracy with 84.77% while KNN and DT is just 82.25% and 77.06%, respectively. Furthermore, Precision, Recall and F1 Score are also described in the **Table 1.** and **Fig. 6** shows the confusion matrix of these three classifiers. In confusion matrix, it is better to have more instances in the main diagonal (two darker grid), and fewer instances in the remaning two grids. Moreover, **Fig. 7** also shows the receiver operating characteristic (ROC) Curves of these classifiers. It is recommended that the AUC area should have value to be above 0.5. Also, **Fig. 7** shows that the AUC score of LR is higher than KNN and DT, which are 0.6231, 0.5665 and 0.5401, respectively.

4.2. Experimental Results of Proposed Method

In this paper, experimental results are shown and GWO-based feature selection was applied with XGBoost to limit unnecessary features and enhance the performance. **Fig. 8.** shows that fitness function

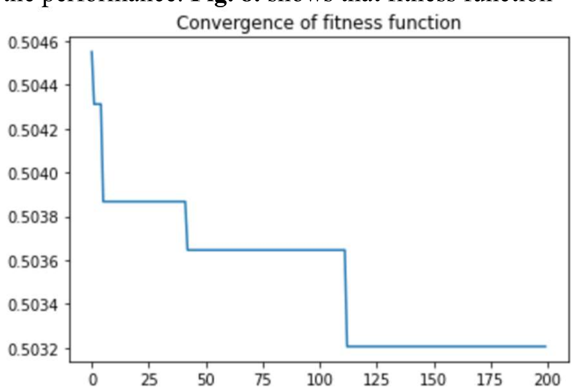


Figure. 8. Convergence characteristics of fitness function

Table 2. Best parameter with hyperparameter tuning

Parameters	Values after tuned
max_depth	3
min_child_weight	3
gamma	0.3
subsample	0.8
colsample_bytree	0.8
reg_alpha	1
learning_rate	0.1

values corresponding to 200 iterations. As we can see, the number of iterations increase, the optimization function decreases until it ends its iteration process, and the best fitness value is 0.50321.

As can be seen, at first, a set of features is selected, and put in the training process using XGBoost and then come to a temporary result. The new subset is selected in the next iteration and continues the process until 50 iterations are reached. Finally, the final subset of 15 features is selected, which is: ‘X1’, ‘X3’, ‘X5’, ‘X6’, ‘X7’, ‘X9’, ‘X11’, ‘X12’, ‘X13’, ‘X14’, ‘X16’, ‘X17’, ‘X18’, ‘X20’ and ‘X22’.

After being tuned, a set of hyperparameters are selected, which enhances the performance of the model. This parameter set is clearly described in **Table 2.** The feature importance graph is shown in **Fig. 9**, respectively.

4.3. Performance Evaluation

As shown in **Table. 2** and **Fig. 7**, the optimized XGBoost-based classifier outperforms others, gaining 85.58% accuracy and 0.6736 AUC score (also higher than the result before proposed framework is implemented, which is 84.77% and 0.6412, respectively). This proves the robustness of the XGBoost classifier together with the hyperparameter tuning technique. **Fig. 10** and **Fig. 11** show the

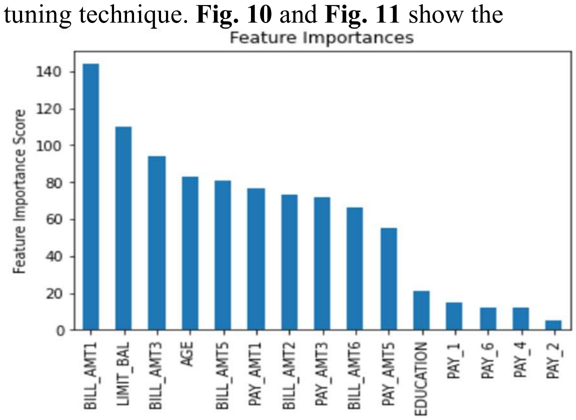


Figure. 9. Feature Importance with 15 selected features

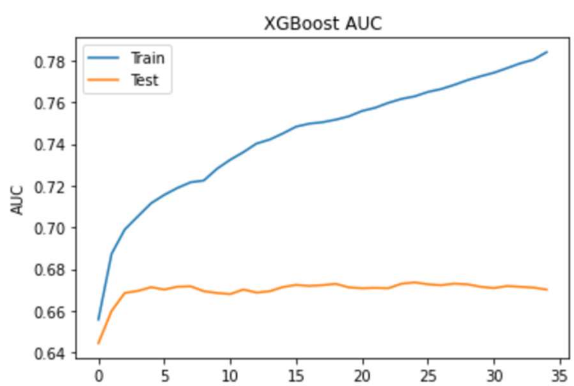


Figure. 10. The train and test AUC score of models

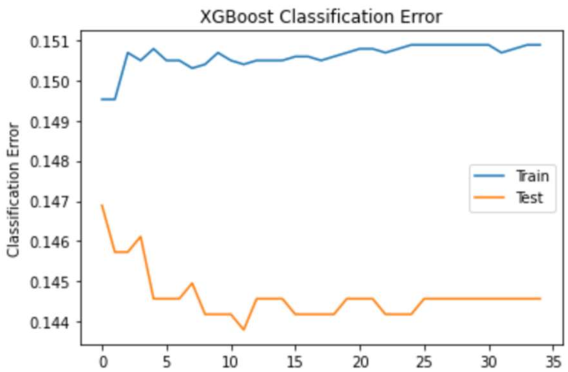


Figure. 11. The train and test classification error of model

model's AUC score and classification error after 200 iterations. The early stopping rounds helps avoid overfitting problem. The best results can be obtained after 33 rounds.

To explain this, DT and KNN are just basic, and common algorithms, and do not have good performance due to overfitting. Next, LR is non-linear and has better performance than the previous two. However, XGBoost is an upper level of tree-based algorithm learning from previous trees and number of its classification trees is bigger than DT. Hence, XGBoost achieves a better performance than all three.

5. Conclusion

In this paper, we have proposed XGBoost classifier together with BGWO metaheuristic algorithm for feature selection. Proposed model is then compared with other three conventional ML algorithms. This feature selection techniques has good performance in selecting relevant features, and eliminate redundant and irrelevant ones. The computational results indicate that fewer features are chosen and better performance is achieved. Since the chosen dataset has no missing values, in the future, other real-world datasets which contain anomalous and noisy values will be applied to further investigate the robustness of the proposed method. Moreover, we aim to improve the exploration and exploitation of BGWO by applying new transfer function, which is introduced by Hu et al. (2020).

Reference

Chen T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System.

El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. *International Conference on Engineering & MIS (ICEMIS).* IEEE.

Emary, E., Zawbaa, H.M., Grosan, C., Hassenian, A.E. (2015), Feature subset selection approach by gray-wolf optimization. *Afro-European Conference for Industrial Advancement,* 1–13.

Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, **172**, 371–381.

Galindo, J., Tamayo, P. (2000), Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. *Computational Economics*, **15**, 107–143

Guo, J., Yang, L., Bie, R., Yu, J., Gao, Y., Shen, Y., & Kos, A. (2019). An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring. *Computer Networks*, **151**, 166– 180.

Hossin M. and Sulaiman M.N., (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, **5**(2), 01–11.

Hu, P., Pan, J.-S., & Chu, S.-C. (2020). Improved Binary Grey Wolf Optimizer and Its application for feature selection. *Knowledge-Based Systems*, **195**(105746), 105746.

Jain A. and Zongker D. (1997). Feature selection: evaluation, application, and small sample performance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(2), 153–158.

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, **40**(13), 5125–5131.

Li H., Cao Y., Li S., Zhao J. and Sun Y., (2020) XGBoost Model and Its Application to Personal Credit Evaluation, *in IEEE Intelligent Systems,* **35**(3), pp. 52-61.

Li, S., & Zhang, X. (2020). Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm. *Neural Computing & Applications*, **32**(7), 1971–1979.

Mirjalili S., Mirjalili S. M., and Lewis A., (2014) Grey Wolf Optimizer, *Advances in Engineering Software,* **69**, 46–6.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics,* 7.

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. (Kouros). (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention,* **136**, 105405.

Ryu, S.-E., Shin, D.-H., & Chung, K. (2020). Prediction model of dementia risk based on XGBoost using derived variable extraction and hyper parameter optimization. *IEEE Access: Practical Innovations, Open Solutions*, **8**, 177708–177720.

Too J., Abdullah A. R., Mohd Saad N., Mohd Ali N., and Tee W., (2018), A New Competitive Binary Grey Wolf Optimizer to Solve the Feature Selection Problem in EMG Signals Classification, *Computers*, **7**(4).

Twala, B. (2010). Multiple classifier application to credit risk assessment. *Expert Systems with Applications,* **37**(4)*,* 3326–3336.

Yan K., & Zhang D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, **212**, 353– 363.

Ye, Q. Y., Rao, H., & Ji, M. S. (2017). Business sales forecast based on XGBOOST. *Journal of Nanchang University*, **3**, 275–281.

Yu L. and Liu H. (2004), Efficient Feature Selection via Analysis of Relevance and Redundancy, *J. Mach. Learn. Res*., **5**, 1205–1224.

Zhang, Y., Wang, S., Phillips, P., & Ji, G. (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowledge-Based Systems*, **64**, 22–31.

Zorarpacı E. and S. A. Özel, (2016), A hybrid approach of differential evolution and artificial bee colony for feature selection, *Expert Systems with Applications,* **62**, 91–103.

APPLICATION OF SIMULATION IN HEALTHCARE SYSTEM: A CASE STUDY OF OUTPATIENT CLINIC IN A HOSPITAL IN HO CHI MINH CITY

Nguyen Truong Nguyen[1], Doan Le Thao Vy[2], Le Quang Trung[3], Nguyen Vo Cong Thanh[4] and Nguyen Bui Thanh Truc[5]

[1] *School of Industrial Engineering and Management, International University – VNU HCMC, ntnguyen@hcmiu.edu.vn*

[2] *School of Industrial Engineering and Management, International University – VNU HCMC, doanlethaovy@gmail.com*

[3] *School of Industrial Engineering and Management, International University – VNU HCMC, trunglequang1010@gmail.com*

[4] *School of Industrial Engineering and Management, International University – VNU HCMC, smathanhh@gmail.com*

[5] *School of Industrial Engineering and Management, International University – VNU HCMC, nbttruc94@gmail.com*

Abstract. Healthcare is one of the highest-growing industries in the world. However, despite annually significant financial investments, healthcare systems have been facing the problem of overcrowding. Staff and facilities in many systems are required to operate intensively to provide enough health service. The situation is deteriorating during this time of pandemic, especially for outpatients. Deficient capacity results in prolonged patients' wasting time and negative impact on their health treatment. To address this issue, a discrete event simulation model is developed utilizing Arena simulation software and real-time data collected from a major hospital in Vietnam. Based on the simulation model that had been validated by hypothesis testing, we work out the system's downsides in procedure flow and resource utilization then propose alternatives to eliminate process delays as well as deploy the system to its best. The new system performance is evaluated based on patients' waiting time and the system's operational efficiency according to regulations by the Ministry of Health.

Keywords: healthcare system; simulation; Arena; hypothesis testing; waiting time

1. Introduction

Health care is one of the government's primary priorities, with a long tradition of creation and development, as well as the nation's heroic past, the whole health sector will continually strive and rise up to contribute to the future. Make a sacrifice, serve your nation, and serve the people. During more than 40 years of innovation and development, Vietnam's health sector has achieved many great achievements, attracting special attention from the international community. In 2015, the whole health system examined and treated more than 146 million patients, an increase of 4.5% (approximately 6.3 million) compared to 2014. In which, the number of inpatients was approximately 13.5 million people, the number of hospital beds per person established at three central, provincial, and district levels (including private and medical hospitals) is 32.1, an increase of 7.4 beds per thousand people. compared to 2012 (24.7 beds/thousand people). After a short period of time, spectacular results were achieved on the deployment of health insurance to over 70% of the population participating in health insurance in 2013.

In 2020, the world witnessed the raging of the Covid-19 pandemic, which infected more than 105 million people, claimed the lives of more than 2.3 million people, and affected all aspects of human life. In Vietnam, with the drastic and correct direction of the Party and Government, the participation of the National Steering Committee, the Ministry of Health, and other ministries, departments, branches, and localities, especially the consensus. With the wisdom and active participation of all classes of people, we have been proactive in controlling and controlling outbreaks. In particular, the health sector has promoted the tradition of solidarity, creativity, is not afraid of difficulties and hardships, is ready to stand firm in all hot spots, has succeeded in controlling outbreaks, minimizing the number of people. morbidity and mortality as well as the extent of disease transmission. Besides contributing to the success in the prevention and control of the Covid-19 pandemic, with innovative thinking, the health sector also has many breakthroughs to better serve the people. The introduction of a remote medical examination and treatment system connecting 1,500 hospitals to help people enjoy medical services at the grassroots level; 97.5 million people's health records were established, creating a premise to build smart healthcare... The whole industry is actively implementing digital transformation with the introduction of a series of digital health platforms, applications, and technology. declare prices of drugs, equipment, medical services, etc. to create favourable conditions for people and businesses. However, there are certain difficulties that must be addressed.

Overcrowding is especially acute in the national hospital, which has a team of experienced doctors and medical equipment that has been modernized. Many critical concerns, such as (1) congestion, (2) medical procedures, (3) binding rules (Ministry of Health), and (4) the efficacy of system management information, have been addressed and resolved. In this study, discrete event simulation (DES) is used to simulate operations in outpatient departments, which are impacted by overpopulation. The main objective of this study was to create simulation models for the hospitals under investigation. Through the data and results obtained, we proceed to use simulation, identify bottlenecks and possible difficulties in current operations. Since then, a number of proposed measures to improve the operational efficiency of the hospital compared to the current reality were carried out and changed to help improve service quality.

2. Literature Review

Solving the hospital facility layout problem has become one of the most interesting topics of operations research worldwide. The method of simulation optimization based on metaheuristic algorithms for handling the facility layout problem has been developed recently. However, the hospital facility layout problem is a new research in Viet Nam, there has been little application in this field. Therefore, there is an opportunity to apply the method to solve the hospital facility layout problem in Vietnam.

This paper briefly proposes a basic design and development of a DES model, which mainly relied on the object oriented paradigm (OOP) within a physician network. Furthermore, this study demonstrates a visual simulation environment to help illustrating and communicating the findings as well as the full development process, which includes detailed descriptions like data collection, modeling approaches and key performance indicator of hospital. This will enable us to deeply understand the mechanisms of a physician clinic, which will enhance the theoretical factual foundations for making a firm decision (Swisher et al., 2001).

By analyzing the patients' behavior in hospital, this study will solve two kinds of resources to optimize: the human resources and the equipment such as patient beds to determine the best allocation of resources and its applications in hospital. By implementing the outcome of the study and optimization of the hospital resources, waiting time for the patients could be reduced significantly, and also the related costs can be controlled properly.

Calgary Laboratory Services located in Canada has investigated about a network of outpatient clinics with the purpose of determining the impact on patient demand. It shows that the number of laboratory facilities would be reduced from 25 to 18, 12 or 6 facilities, respectively. The development of isolated event simulation model is operated to anticipate the profit of pooling and to illustrate the optimum service provision. There were 18 amenities suggested, all of which have been implemented in practice. The execution has the perfectly unexpected effect on patient demand that system dynamics indicate. Finally, that system is credited with contributing significantly to developing the simulation model in order to define the overall system behaviours (Rohleder et al., 2007).

Additionally, another simulation model is to check out the various appointment schedules was conducted in the Ear, Nose, Throat (ENT) outpatient department near London. This methodology was based on feedback from consultants and experienced high-level staff, who revealed that if clinics started on time, 15 minutes of waiting time could be reduced on average. Furthermore, a detailed analysis of the schedules reveals that arranging patients into large blocks should be avoided in order to save an additional 8 minutes in patient wait time (Harper et al., 2003).

The next research conducted at a local hospital in Taiwan – Chiavi, which put all of the focus on the usage of simulation to cut off the out-patient queues of a dermatology outpatient department. By using "what-if" scenarios for simulation depict a considerable growth in performance if they add an extra session on Monday afternoon. The length of stay is decreased by 47% after implementation; only 3% of patients have a LoS of more than 1.5 hours, the maximum queue length is reduced to a third of its original value, and physician utilization is reduced by 78% (Huarng et al., 1996).

The results of the next study were used to test four different scheduling strategies at the Gero prefectural in Gifu, Japan, utilizing DES to test B2 (Baily), Rising, 15MIN, and SPTBEG. Finally, the results show that the SPTBEG is best for minimizing waiting time, while the 15MIN rule is the best for reducing physician idle time. The combination of a hybrid 15 MIN rising therefore would be most suitable (Wijewickrama et al., 2006).

There are 2 beneficial influences on applying simulation in outpatient departments generally:

1) Gain a better understanding of the mechanisms operating within complex systems.

2) Perform "what-if" scenarios on the model without affecting with the current running system. Scenarios can be used to determine the best performing schedule rule, for example (Wijewickrama, 2006). Due to the ease of application as an optimization approach, other published studies make wide use of 'what-if'-scenarios, especially when combined with visual interaction (Swisher et al., 2001). When it comes to planning or implementing health-care facilities, simulation became crucial. Downsizing while still providing vital services is an issue that Rohleder et al. (2007) has resolved by evaluating the amount of laboratory facilities to serve the region.

3. Simulation practical for outpatient treatment (Selection of case study: Hospital X at Ho Chi Minh City)

In Ho Chi Minh city, due to overcrowding of patients seeking medical care system, regular medical care is overloaded. Therefore, the normal analysis methods cannot handle the complexity well. In that case, the simulation method is applied to study the medical area of a large number of hospitals in Ho Chi Minh City. By applying simulation program development and research system, this study will solve the bottlenecks and improves the system of outpatient clinics in the hospital X based on ARENA simulation software.

3.1 The process of medical examination

The medical examination and treatment process depends on the type of disease, the examination process of patients with health insurance and no health insurance does not differ from the stage of examination and sub-clinical design. The counter will charge a different examination fee depending on the property with or without insurance. Patients who are not sub-clinically fee prescribed will receive medicines at the hospital medicine counter if they have health insurance or go to buy medicines at the pharmacy counter. For patients who are prescribed sub-clinically, after finishing the procedure will return to the original clinic for the doctor to diagnose and prescribe the medicine.

The figure below depicts the standard examination procedure for a general hospital examination ward. For patients with health insurance, it is not necessary to pay the initial deposit, but to come directly to collect the order number from the clinic. After completing the clinical examination and returning to the originating clinic, the insured patient must follow the health insurance prescription approval procedures, pay the inspection fee, and seek medical treatment. For non-health-insured patients, after the clinical examination and return to the original clinic, they come to buy medicine at the hospital pharmacy.

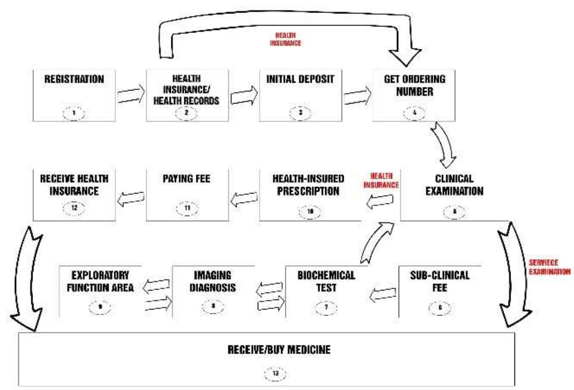
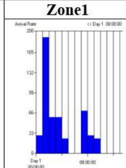
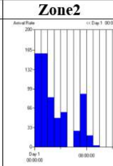
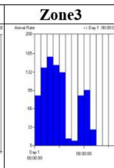


Figure 1: A general patient flow at the hospital

3.2 Data collection and Input analysis

One data set of this simulation model is the distribution of patient arrival time in 10 continuous time slots, from 06:00 AM to 04:00 PM. The system is disaggregated into three different zones, thus patient arrivals are studied independently for each zone. Input data also includes the distribution of system's processing time through the operational process. It can be implied that every patient yields a random time of arrival. In addition, depending on each type of disease and illness will every patient have different time of examination. Data are collected from direct observation by the assistance of IT department in the hospital.

Table 1: Data collection

Time Slot	Zone1	Zone2	Zone3	Arrival number patients (Direct Observation)
				
6:00-7:00	30	160	90	**280**
7:00-8:00	190	160	140	**490**
8:00-9:00	60	85	160	**305**
9:00-10:00	60	50	145	**255**
10:00-11:00	25	60	132	**217**
11:00-12:00	0	0	13	**13**
12:00-13:00	0	28	9	**37**
13:00-14:00	70	91	90	**251**
14:00-15:00	30	20	100	**150**
15:00-16:00	25	3	29	**57**
Total	**490**	**657**	**908**	**2055**

From the table above, it can be seen that there is a total of over 2000 patients coming to the hospital for a check-up. The patient volume is highest during the time between 07:00 AM and to 08:00 AM, amounts to 490 people. On the other hand, this number hits the bottom at 13 patients during the noon time from 11:00 AM to 12:00 PM. Of all the three zones, zone 3 has the most total arrival of patients with 908 people, whereas zone 1 has the least total patients visiting. Collected data will be inspected and analyzed for the most appropriate parameters of the distribution to be selected for the model.

3.3 Simulation model (Arena logic)

The figure below illustrates the simulation of hospital's layout and our simulation logics. Notation for each number symbol is listed in the table beside. The simulation logics express 10 procedures of the system's process. First, the logic of patient arrival at zone 1, 2 and 3 is created by the input of collected data. Patients are then simulated to queue up for registration at the counter area and proceed to the examination room area. Other logics include the process of patients taking blood test, being diagnosed, receiving result and taking medicine at the hospital pharmacy. We also develop the logics of the procedure in exploratory function area and the assignment of the rate of diseases. As for the layout, the whole system is simulated through entrance zones and three other specialized areas.

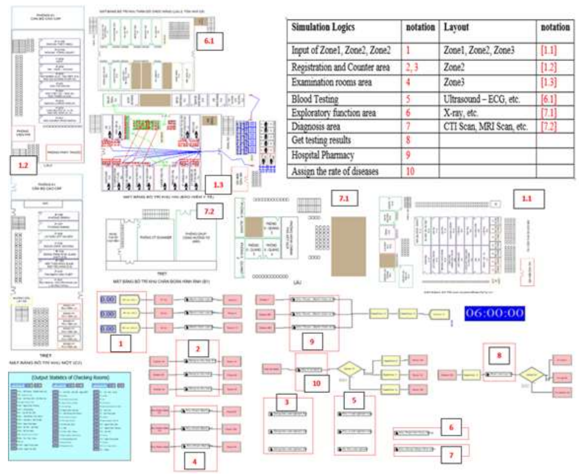


Figure 2: Simulation model

3.4 Key performance indicator of hospital

- ***KPI of para-clinical equipment***

$$U_i = \frac{\text{Time for use the equipment}}{\text{Total time}} \qquad (1)$$

U_i: efficiency preclinical equipment which $i = 1,2,3,4, .., n$ is devices

- ***The evaluate indicator of clinic***

Based on Decision No. 1313 / QD - BYT on 04/22/2013 of the Minister of Health, the average maximum strives examination room for health check 50 patient / 8 hours. Each clinic has indicators on the number patients of days N proposed as follows:

$$\text{EIC} = \begin{cases} 0, \text{ if } N \leq 50 \\ \frac{N - (50\times c)}{50\times c}, \text{ if } N > 50 \end{cases} \qquad (2)$$

EIC: the evaluate indicator of the clinic
N: the number of patients in each clinic
c: the number of doctors in each clinic

EIC indicators show the status, performance of the clinic. If the EIC = 0 ($N \leq 50$ patients), the clinic meets the medical examination of the patient. Conversely, if the EIC > 0 that is the clinic has reached the maximum level. A higher EIC represents overcrowding of the clinic.

- ***Waiting time indicators of the patient***

A matrix defined with X_{nm} as the elements indicating the waiting time of patient *n* for activity *m*

Average waiting time

$$= \frac{\sum_m X_1 + \sum_m X_2 + \sum_m X_n}{n} \quad (3)$$

- ***Service quality indicators for patients***

Service quality indicators for patients (IBN) depends on medical examination time, waiting time and moving time represented by the formula.

$$\text{IBN} = \frac{\text{PT}}{\text{WT} + \text{MT}} \quad (4)$$

PT: the processing time of patients
WT: waiting tine
MT: moving time
IBN index higher patient satisfaction while increasing medical examination. Patient satisfaction when examination time (PT) much, doctor diagnosed carefully and more accurate results. The waiting time for the patient and move less, less waiting time will make patients more satisfied it to the examinations in the hospital.

3.5 Verification and Validation

Verification is a process that the modeler to ensure that the model works as intended. To verify and debug models, Arena's Run Controller tool is used. Model validation is the task that ensures the model works in the same way as the real system. It also ensures that the results it produces are within an acceptable level of accuracy. The simulation model was run for 10 statistically independent replications, as we did in our study. The simulation time periods from 6:00 to 17:00 to compare the model and the actual data.

The two-sample t-test for unpaired data is defined as:

$H_0 : \mu_d = 0$

$H_1 : \mu_d \neq 0$

The test statistics is:

$$t_0 = \frac{\overline{Y_1} - \overline{Y_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \quad (5)$$

Where $\overline{Y_1}$ and $\overline{Y_2}$ are the sample means, $s_1{}^2$ and $s_2{}^2$ are the sample variances.

Significant level: $\alpha = 0.05$

4. Results

The process of running the simulation, simulation programs to the actual situation of the patient movement in health care system based on the logic model has been built.

ARENA models are used to simulate the hospital operations. The simulated hospitals will operate from 6:00am to 5:00pm. For each scenario (of a hospital and a specific policy), 30 independent replications are performed, and the recorded information is used to evaluate the system. Several measures are used to assess the performance of hospitals:

- Throughput (the number of patients out of the system)
- Waiting times of patients
- Total moving times of patients
- Utilization of each function room
- The evaluate indicator of clinic
- Service quality indicators for patients

4.1 Performance of the current systems

The results of a simulation run for a day at the hospital were recorded in the table. In a typical day, the number of patients go to the hospital is approximately 2158, while the number of patients examined and left is approximately 1834. Furthermore, the average time at the hospital, that is, the time the patients recorded when they arrived and when they left, is roughly 4.56 hours; additionally, the average moving time is 0.17 hour, and the average waiting time for patients is approximately 3.95 hours.

Table 2: Parameters of the current system

No.	Parameter System	Current situation in the hospital X
1	Patients in	2158
2	Patients out	1834
3	The average time in the hospital	4.56 hour
4	The average moving time	0.17 hour
5	The average waiting time	3.95 hour

4.2 The waiting time of patients

The waiting time of patients in 10 replications is show in the flowing table.

Table 3: The waiting time of patients

The waiting time	The number of patients
	1834
t < 0.5	202 ~ 11.01%
0.5 ≤ t < 1	147 ~ 8.02%
1 ≤ t < 1.5	257 ~ 14.01%
1.5 ≤ t < 2	293 ~ 15.98%
2 ≤ t < 2.5	248 ~ 13.52%
2.5 ≤ t < 3	183 ~ 9.98%
3 ≤ t < 3.5	128 ~ 6.98%
t ≥ 3.5	367 ~ 20.01%

For 10 replications, the overall processing time is approximately 4908.89 hours, and the total waiting time is 38175.7 hours. The ratio between each is 7.77. This ratio is quite high, showing that the patient wastes a significant amount of time waiting for medical examination, while the examination duration is less than several times.

Because many patients visit and depending on the proportion of patients, we will have a total of 36 types of preclinical and clinical. The chart below will show the distribution time of 36 types of medical examination and treatment at hospital X.

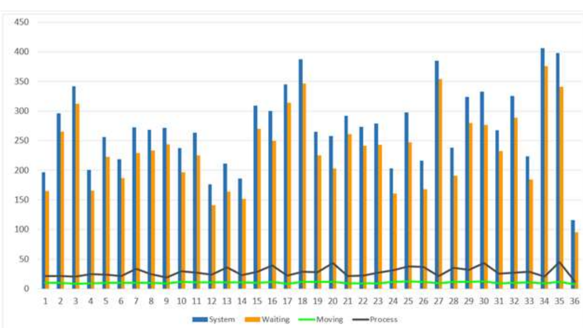


Figure 3: Distribution of 36 types of medical examination and treatment at hospital X

4.3 Service quality indicators for patients

An evaluation index clinic "EIC". On average, each examination rooms only strives maximum 50 patient examination/8 hours/ doctor. In this case, we assume doctors for all clinics is 1. EIC indicators show situation, performance of the clinic. If indicators EIC = 0 (N <= 50 patients), the clinic medical working good. Conversely, if EIC > 0 (N> 50), i.e. the clinic has reached the maximum level. EIC expressed higher the overload in the clinic.

Table 4: EIC index of function rooms at zone 1

No	The Function rooms		Patients	EIC index	Situation
ZONE1					
1	P101	Internal Medicine	26	0	
2	P102	Interventional Cardiology	50	0	
3	P103	Neurology Internal - Gastroenterology	35	0	
4	P104	Neurology Surgical – Traditional Medicine	53	0.06	Overload
5	P105	Traumatic Surgical	21	0	
7	P107	Dentist	58	0.16	Overload
8	A1	Senior Staff Clinic	25	0	
9	P201	Internal Medicine	27	0	
10	P202	Respiratory Medicine - Oncology	42	0	
11	P203	Thoracic Surgery	13	0	
12	P204	Endocrinology - Nephrology	62	0.24	Overload
13	P205	Cardiology Internal	43	0	
14	P206	Infection - Endocrinology – Musculoskeletal Internal	54	0.08	Overload
15	P208	Otorhinolaryngology	28	0	
16	P209	Ophthalmology	76	0.52	Overload
17	P210	Surgical Urology - P211	13	0	
		General Surgery - P212	31	0	

Table 5: EIC index of function rooms at zone 2

ZONE2					
1	P1	Endocrinology – Musculoskeletal Internal - Gastroenterology	68	0.36	Overload
2	P2	Internal Medicine	68	0.36	Overload
3	P3	Neurology Internal - Gastroenterology	74	0.48	Overload
4	P4	Cardiology Surgical	45	0	
5	P5	Cardiology Internal	70	0.4	Overload
6	P6	Traumatic Surgical	140	1.8	Overload
7	P7	Internal Medicine	55	0.1	Overload
8	P8	Cardiology Internal	50	0	
9	P9	Neurology Surgical – Traditional Medicine	35	0	
10	P10	Obstetrics and Gynecology	50	0	
11	P11	Ophthalmology	47	0	
12	P12	Otorhinolaryngology	68	0.36	Overload
13	P13	General Surgery - Urology	31	0	
14	P14	Pyretic	0	0	
15	P15	Infection	0	0	
16	P16	Dermatology	41	0	

Table 6: EIC index of function rooms at zone 3

ZONE3					
1	P1	Otorhinolaryngology	56	0.12	Overload
2	P2	Cardiology	60	0.2	Overload
3	P3	Ophthalmology	19	0	
4	P4	Neurology Internal	64	0.28	Overload
5	P5	Health care services for working	No data		
6	P6	Gastroenterology	31	0	
7	P7	Internal Medicine	68	1.36	Overload
8	P8	Cardiology	40	0	
9	P9	Thoracic Surgery	28	0	
10	P10	Neurology Surgical	30	0	
11	P11	Thoracic Surgery	40	0	
12	P12	Endocrinology	No data		
13	P14	Respiratory	17	0	
14	P15	General Surgery	21	0	
15	P17	Obstetrics and Gynecology	10	0	
16	P18	Lithotripsy –Urology Surgery	7	0	
17	P21	Dentist	8	0	

4.4 Proposed solution

At the entrance zone, the patient must pay a fixed upfront cost (according to the hospital's policy) and receive the ordinal number of the clinic. After successful registration, the patient will be present at the preliminary examination clinic, then conducting order tests (for patients who need testing) or being prescribed a prescription (for patients who did not test or who returned to the trial design). Once there is an order for testing, the patient will keep the order sheet and go to the respective laboratories to perform. After the test results and prescription are available, the patient will pay the fee, if there is more than the paid fee, the patient must pay more; otherwise, a balance refund will be eventually settled. After the payment, the patient will receive the medicine from the pharmacy and the results from the doctors and then exist. In the new patient flow, prepayment is applied to check-in and the information system is combined to improve communication between different functions. In the new testing rule, patients must follow each step to implement this rule.

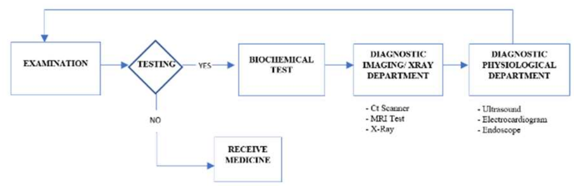


Figure 4: The suggested process flow for the hospital

After getting the results of current situation hospital X, we began looking for ways to improve processes to reduce waiting time, travel time and utilization of the functions in the hospital room. The results are shown as follows:

Table 7: Wait time of current system

Wait Time (Current system)						
	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ZONE1	86.0421	5.57	72.5204	99.38	0.00	498.53
ZONE2	103.55	5.21	92.5709	115.20	0.00	797.62
ZONE3	119.31	5.92	107.78	133.37	0.00	549.68

Table 8: Wait time of improved system

Wait Time (Improve System)						
	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ZONE1	82.0590	3.55	76.2595	92.4721	0.00	497.83
ZONE2	94.2897	7.36	80.5533	107.65	0.00	535.15
ZONE3	114.15	4.27	105.19	123.03	0.00	570.56

As we have seen, after improvement, the average waiting time of patients was reduced. For example, in the Zone 1, the maximum waiting time before improvement is 797 minutes, but the time was reduced to 535 minutes after improved.

4.5 Further discussion

This study proves that simulation provides a straightforward and cost-effective approach for us to evaluate the healthcare system' performance under various alternatives and identify its bottlenecks. To improve the complicate and sluggish system in this study, it is highly recommended to simplify the operational flow and develop a management information system. The introduced management information system should have a strong hardware and a big data storage.

Moreover, regarding the excessively high utilization of the hospital's resources and the deficiency of patient services, it is of great necessity for the system to expand its capacity and allocate facilities reasonably. Spreading the workload of these major hospitals to local hospitals is also a significant macro strategy that will address the problem of overcrowding in long term.

Another root cause for the system's underperformance is the absence of an effective appointment scheduling system. Within a limited capacity, unplanned patient arrivals for non-emergency check-ups, or late arrivals, or patient no-shows can force patients to queue up from early in the morning till late in the afternoon. Prolonged waiting time not only upsets patients but also drains health staffs. Therefore, the investment of an appointment scheduling system can help distribute patient arrivals appropriately to lessen the issue of overcrowding as well as increase overall productivity of the hospital.

5. Conclusion

Simulation is a useful tool to help us control changes, reduce risks, analyse significant data, and improve hospital performance by optimizing existing resources. In this study, a simulation model of two large hospitals in Ho Chi Minh City, Vietnam is built, and the change is analysed. The simulation results show that the simulated object has many problems and bottlenecks in the current system. In addition, the simulation models also show a deep concern about the long average waiting time of patients for care and use of medical services, which affects the effectiveness of medical care and the quality of patient experience.

Overall, the subject met its objectives. Using computational modeling tools and research data on the condition of medical examination systems at several big hospitals in Ho Chi Minh City are collected. A data analysis phase is carried out and preserved as the foundation for future modifications. Centralizing information and process data rapidly so that hospital administrators can understand the situation at clinics in real-time. As a foundation for the development and enhancement of service utilization quality. A model was examined and presented to tackle the problem of improving the health system to decrease examiner overload with a scarcity of medical staff and optimize medical treatment efficiency. by considering the aspect of minimizing patient waiting time.

Reference

Swisher, J. R., Jacobson, S. H., Jun, J. B. and Balci, O. (2001). Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation. Computers and Operations Research, 28, 105-125.

Rohleder, T., B, D. & Baskin, L. (2007). Modeling patient service centers with simulation and system dynamics. Health Care Management Science,10, 1-12.

Harper, P. R. & Gamlin, H. M. (2003). Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. OR Spectrum, 25, 207-222.

Huarng, F. & Lee, M. H. (1996). Using simulation in out-patient queues: a case study. International Journal of Health Care Quality Assurance, 9, 21-25.

Wijewickrama, A. K. A. (2006). Simulation analysis for reducing queues in mixed patients' outpatient department. International Journal of Simulation Modelling, 5, 56-68.

Feili, H.R. (2013). "Improving the Health Care Systems Performance by Simulation Optimization." Mathematics and Computer Science, 7, 73-79.

Michael C.Fu, F.W.G. (2005). Simulation optimization A review, new developments, and applications. Winter Simulation Conference.

Yang, M., et al. (2013). The call for equity: simulation optimization models to minimize the range of waiting times. IIE Transactions, 45(7), 781-795.

S. Jannat, A. A. K., and Sanjoy, K.P (2010). Optimal Solution for Multi-Objective Facility Layout Problem. International Conference on Industrial Engineering and Operations Management, 9-10.

M. C. Fu. (2002). Optimization for simulation: Theory vs. practice. INFORMS J. on Computing, 14 (3), 192–215.

Nguyen, T.N., Phong, H.T., Quoc, H.D. (2017). Simulation Application in Healthcare Services: A Case Study of the Outpatient Clinic in the Hospital.

Appendix

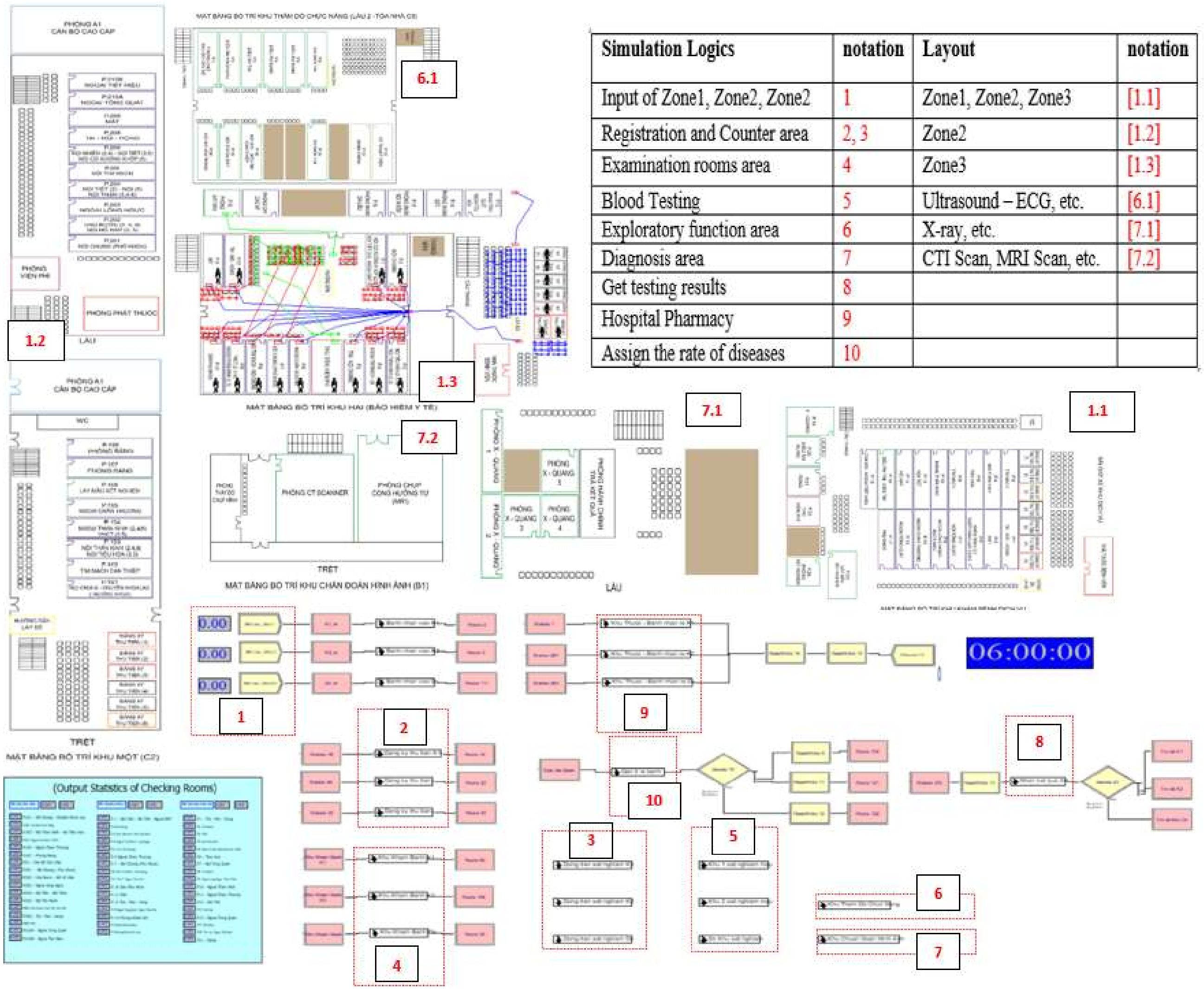


Figure 2: Simulation model

Table 9: The distribution of processing time in the hospital X

No	The Function rooms		The distribution	Unit
ZONE 1				
1	P101	Internal Medicine	UNIF(6,7)	Min
2	P102	Interventional Cardiology	UNIF(5,7)	Min
3	P103	Neurology Internal - Gastroenterology	UNIF(4,6)	Min
4	P104	Neurology Surgical – Traditional Medicine	UNIF(3.5,5)	Min
5	P105	Traumatic Surgical	UNIF(6,7)	Min
6	P107	Dentist	UNIF(4,6)	Min
7	A1	Senior Staff Clinic	UNIF(4,6)	Min
8	P201	Internal Medicine	UNIF(6,7)	Min
9	P202	Respiratory Medicine - Oncology	UNIF(6,7)	Min
10	P203	Thoracic Surgery	UNIF(3.5,5)	Min
11	P204	Endocrinology - Nephrology	UNIF(6,7)	Min
12	P205	Cardiology Internal	UNIF(6,7)	Min
13	P206	Infection - Endocrinology – Musculoskeletal Internal	UNIF(3.5,5)	Min
14	P208	Otorhinolaryngology	UNIF(6,7)	Min
15	P209	Ophthalmology	UNIF(4,6)	Min
16	P210	Surgical Urology - P211	UNIF(3.5,5)	Min
		General Surgery - P212	UNIF(10,12)	Min
ZONE 2				
1	P1	Endocrinology – Musculoskeletal Internal - Gastroenterology	UNIF(6,7)	Min
2	P2	Internal Medicine	UNIF(5,7)	Min
3	P3	Neurology Internal - Gastroenterology	UNIF(4,6)	Min
4	P4	Cardiology Surgical	UNIF(3.5,5)	Min
5	P5	Cardiology Internal	UNIF(3.5,7.5)	Min
6	P6	Traumatic Surgical	UNIF(4,6)	Min
7	P7	Internal Medicine	UNIF(4,6)	Min

No	The Function rooms		The distribution	Unit
8	P8	Cardiology Internal	UNIF(6,7)	Min
9	P9	Neurology Surgical – Traditional Medicine	UNIF(6,7)	Min
10	P10	Obstetrics and Gynecology	UNIF(3.5,5)	Min
11	P11	Ophthalmology	UNIF(6,7)	Min
12	P12	Otorhinolaryngology	UNIF(6,7)	Min
13	P13	General Surgery - Urology	UNIF(3.5,5)	Min
14	P14	Pyretic	UNIF(6,7)	Min
15	P15	Infection	UNIF(4,6)	Min
16	P16	Dermatology	UNIF(3.5,5)	Min
MEDICAL SERVICES ON REQUEST ZONE				
1	P1	Otorhinolaryngology	UNIF(6,7)	Min
2	P2	Cardiology	UNIF(5,7)	Min
3	P3	Ophthalmology	UNIF(4,6)	Min
4	P4	Neurology Internal	UNIF(3.5,5)	Min
5	P5	Health care services for working	UNIF(6,7)	Min
6	P6	Gastroenterology	UNIF(4,6)	Min
7	P7	Internal Medicine	UNIF(4,6)	Min
8	P8	Cardiology	UNIF(6,7)	Min
9	P9	Thoracic Surgery	UNIF(6,7)	Min
10	P10	Neurology Surgical	UNIF(3.5,5)	Min
11	P11	Thoracic Surgery	UNIF(6,7)	Min
12	P12	Endocrinology	UNIF(6,7)	Min
13	P14	Respiratory	UNIF(3.5,5)	Min
14	P15	General Surgery	UNIF(6,7)	Min
15	P17	Obstetrics and Gynecology	UNIF(4,6)	Min
16	P18	Lithotripsy – Urology Surgery	UNIF(3.5,5)	Min
17	P21	Dentist	UNIF(10,12)	Min

Table 10: The number type of disease in Medical services on request Zone

No	Rooms	Total	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	P1	47				1		1	11			1			2			1									1		1			1						27
2	P2	52	6	1			1			2				1							2																	39
3	P3	18						1		1																												16
4	P4	60	1	2			1	3		1				2																								50
5	P5																																					
6	P6	35	4			1		2	7					2	1							1																17
7	P7	71	3	1				1		1											2						1		13							2		47
8	P8	38	2	1						1				1							6	1																26
9	P9	30	8			1		2		2				2							1																	14
10	P10	30	6			1		2						5																								16
11	P11	40				1		1						1				1											1									35
12	P12	0																																				
13	P14	16	1					1		1	1			1		1					3																	7
14	P15	18	6						1			1		1	1						2		1										1					4
15	P17	17	8								2				2																							5
16	P18	11	2																																			9
17	P21	3						1																														2

Table 11: The cumulative in Medical services on request Zone

No	Rooms	Total	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	P1	47	0.00	0.00	0.00	0.02	0.02	0.04	0.28	0.28	0.28	0.30	0.30	0.30	0.34	0.34	0.34	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.38	0.38	0.40	0.40	0.40	0.43	0.43	0.43	0.43	0.43	0.43	1.00
2	P2	52	0.12	0.13	0.13	0.13	0.15	0.15	0.15	0.19	0.19	0.19	0.19	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	1.00
3	P3	18	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	1.00
4	P4	60	0.02	0.05	0.05	0.05	0.07	0.12	0.12	0.13	0.13	0.13	0.13	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	1.00
5	P5	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	P6	35	0.11	0.11	0.11	0.14	0.14	0.20	0.40	0.40	0.40	0.40	0.40	0.46	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	1.00
7	P7	71	0.04	0.06	0.06	0.06	0.06	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.11	0.11	0.11	0.11	0.11	0.11	0.13	0.13	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.34	0.34	1.00
8	P8	38	0.05	0.08	0.08	0.08	0.08	0.08	0.08	0.11	0.11	0.11	0.11	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.29	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	1.00
9	P9	30	0.27	0.27	0.27	0.30	0.30	0.37	0.37	0.43	0.43	0.43	0.43	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	1.00
10	P10	30	0.20	0.20	0.20	0.23	0.23	0.30	0.30	0.30	0.30	0.30	0.30	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	1.00
11	P11	40	0.00	0.00	0.00	0.03	0.03	0.05	0.05	0.05	0.05	0.05	0.05	0.08	0.08	0.08	0.08	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	1.00
12	P12	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	P14	16	0.06	0.06	0.06	0.06	0.06	0.13	0.13	0.19	0.25	0.25	0.25	0.31	0.31	0.38	0.38	0.38	0.38	0.38	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	1.00
14	P15	18	0.33	0.33	0.33	0.33	0.33	0.33	0.39	0.39	0.39	0.44	0.44	0.50	0.56	0.56	0.56	0.56	0.56	0.56	0.67	0.67	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.78	0.78	0.78	0.78	0.78	1.00
15	P17	17	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.59	0.59	0.59	0.59	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	1.00
16	P18	11	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	1.00
17	P21	3	0.00	0.00	0.00	0.00	0.00	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1.00

MINIMIZE THE TRAVEL DISTANCE OF THE PICKERS IN THE WAREHOUSE A CASE STUDY OF SCHNEIDER ELECTRIC VIETNAM

Dang Vi Anh[1], *Duong Vo Nhi Anh[2], *Nguyen Hoang Huy[3],

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: [1]dangvianh310599 @gmail.com, [2]dvnanh@hcmiu.edu.vn, [3]nhhuy@hcmiu.edu.vn

Abstract: During the sensitive period of the Covid 19 pandemic, when the economy and other industries have been greatly affected, the fact that the Logistics and Supply Chain industry could have that little impact and still remains its smooth development has raised the curiosity of many factories all around the world. Hence, to solve this problem, the factories need to minimize costs in the warehouse by using an optimal warehouse system. In this article, the optimization model for the order picking method is used to assist the organization in deciding the optimum configuration for the Schneider Electric company's warehouse and eliminating the difficulty indicator ascending. This paper introduces a modern that using the Apriori algorithm to check the frequency of items for assigning in the right location and Mix integer linear programming for reducing picker movement distance in manual warehouses. This case study is carried out by the optimization model and implemented by PYTHON Software, CPLEX Software and it is then compared to previous approaches to have the best solution using real data from the study. The findings show that this modern algorithm can produce better results than heuristic and metaheuristic algorithms and is also capable of providing an exact solution.

Keywords: Order Picking system; storage efficiency; PYTHON software; Apriori algorithm; CPLEX software.

1. Introduction

Firstly, depending on the model, the factories choose the right trend as well as the appropriate development direction to reduce costs and improve productivity. Most manufacturing factories focus on two main trends: catch up with technology 4.0 in the market such as implementing Logistic by Electronic, Logistic with Green improvement, and cloud improvement application to reduce capacity. Blockchain technology, AI, or robotics is widely deployed. The perspective of manufacturers has also become closer, they use smart tools to pack and unloading services from containers, loading and unloading goods in packing and yards; Optimizing warehouse space to minimize inventory for products as well as raw materials.

More than ever, during the sensitive period of the Covid 19 pandemic, when the economy and other industries have been toughly affected, the fact that how Logistics and Supply Chain industry could have that little impact and remains its smooth development has raised the curiosity to many factories all around the world. Hence, to solve this problem, the factories need to minimize costs in the warehouse, by the way, an optimal warehouse system. Warehouse, used by most companies whose main function is is to store goods or goods before transferring them to another location. But there may be some additional storage features to maximize the storage capacity. The monitoring, protection, and security function was almost used as a hierarchy of positions in the warehouse.

In this case study, in the first five months of 2021, the number of items produced by Schneider Electric Manufacturer in Vietnam (SEMV) was higher than half of the total volume of products sold in 2020 (figure. 1). It means that the amount of products sold by the company is not affected by other factors. . That led to a not-so-light effect on the Overall Equipment Effectiveness at production but also created a loss cost due to not meeting the capacity for the line. The Pareto loss table below is the result of performance loss, covered by technical loss and organization loss at the Schneider warehouse.

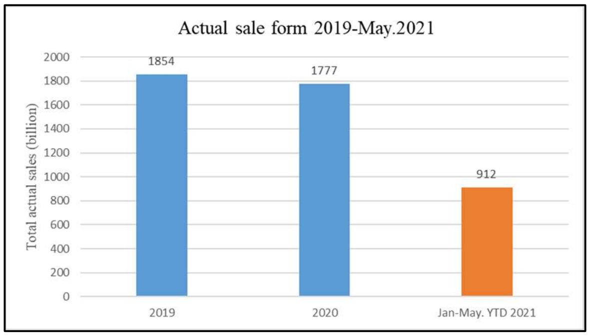


Figure 1. Actual sale of Schneider Electric Company

From 2020 to now, the late picking still happens and takes the lead. The next 2 faults are respectively coming from storage and order picking. The fishbone graphic above (figure. 2) depicts some of the factors that contribute to extended picking times. To begin, two factors are mentioned in the Material, such as improper organization. In terms of humans, the reasons for extended picking processing times include a lack of knowledge about order picking attitude and warehouse worker ability. In terms of machine and technique, there is inadequate hand trolley and long travel distance. After careful examination, the business determined that the major cause for the cost increase is the distribution of products (storage allocation problem). Despite the fact that all of the goods in the present warehouse are placed based on warehouse space, there are still certain inefficiencies in picking.

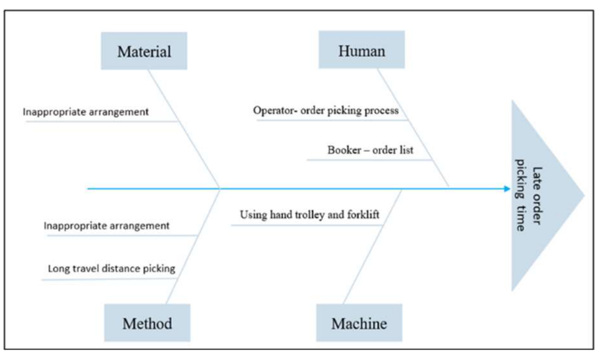


Figure 2. Fishbone diagram of order picking problem.

The main goal of this study is to reduce the total travel distance by picking 20% of pickers to reduce the lost cost of the company. The second aim is to use Excel to test the collected data. In this way, we can recognize irregularities and avoid disrupting the thesis's outcomes. The third goal is to recommend the proposed approach by modifying the existing code with Python to design warehouse layout in order to improve warehouse service quality and reduce operating costs. The Mixed Integer Linear Programming (MILP) formulation is used to solve the order batching problem to evaluate the effectiveness of the new interface. As a result, the purpose of this study is to put in place a model that aids in reorganizing the existing architecture to reduce the picking time for the picking procedure.

2. Literature review

2.1. Warehouse problem:

The key function of a warehouse is to store goods or goods before transferring them to another location. But there may be some additional storage features to maximize the storage capacity. According to studies by Hermann Gruenwald (2015), storage is the most necessary function of storage in the transport center (hub) such as in storage, also referred to as material handling. The monitoring, protection and security function was almost used as a hierarchy of positions in the warehouse.

A warehouse is an important part of the Logistics and supply chain to reduce shipping costs for items, achieve economic efficiency scale in production or purchase or deliver valuable processes and shorten response times. Have a lot of kind Warehouses are categorized into two types: manufacturing warehouses and distribution centers, and their role in the supply chain is known as raw materials, finished goods, distribution, execution warehouse or local warehouse directly to customer needs and added value service warehouses. Although there are differences in warehouse types, as Ramaa, Subramanya, and Rangaswamy (2012) pointed out, most warehouses have the same operating procedure, the difference is made depending on the use needs, basic steps: receive goods; storage; putaway; order picking; repletion; outbound. On the other hand, order choice, accumulation and sorting, packaging goods, cross-fitting and shipping are also mentioned. Emir Zunic and his colleagues (2017) have commented that Warehouse Management Systems (WMS) have played a crucial role in optimizing warehouse logistic operations, archiving industry patterns in supply and demand, and even enabling the handling of products that are close to expiration, out of storage, broken or deposited by consumers. Warehouse operation is built for the purpose of merging the needs and wants of the consumers by efficiently using houses, equipment and labor. (V.Sivakumar and R.Ruthramathi, 2019).

2.2. Datamining:

By James A. Tompkins, Dale A. Harmelink (1994) description, the warehouse storage method includes the transfer of items into their most suitable storage space and is frequently out of storage due to reasons such as fast expansion, seasonal peaks, and sluggish sales. For certain businesses, a warehouse device called to put away will often protect the time when the items travel up to the storage location. Space is a significant feature of the factory. The storage location would be determined and managed according to the following three criteria: how much inventory should be kept in the SKU warehouse; how much and at what time the SKU inventory should be replenished; and where the SKU should be stored in the warehouse and transferred and transported between the various storage areas. The method of storage will directly affect the order picking (Jinxiang Gu and Marc Goetschalckx, 2007).

Datamining: is a relatively recent approach for extracting interesting patterns or rules from huge volumes of data, and it has been utilized in a variety of fields including finance (Anouze, 2011). A comparable concept of support value is proposed by the family grouping policy regarding to the definition of Muchen and his team. According to Andreas Mueller (1995), the area of knowledge discovery in databases, also known as "Data Mining," has grown in popularity in recent years as major corporations recognize the possible importance of the information contained tacitly in their databases. Mining of special, association rules from retail data is a basic data mining activity c. The goal is to identify patterns (or rules) that describe the actions of the purchase's behavior customers from a large database of previous consumer transactions. FIM applies, compute frequency application to stock inventory. Since then, Bhandari (2015), has made the statement one of the most popular algorithms is the Apriori Algorithm, which is used to retrieve frequent itemsets from vast databases and obtain the association law for discovering information. It essentially necessitates two things: minimal funding and minimal trust. First, we search if the items are greater than or equal to the minimum funding, and then we find the frequent itemsets. Secondly, the minimum trust constraint is used to establish association rules. The use of Apriori has saved a lot of time and improved the productivity of the systems because the feature on Apriori is used by the author in the article to improve the efficiency of picking.

2.3. Order Picking Process:

Order Picking has long been mentioned as the most labor-intensive and costly process for virtually any warehouse; the expense of receiving orders is estimated to be as high as 55% of the total operating costs of the warehouse. The selection of a suitable order picking system (OPS) represents one of the key decisions for a company as it has a significant impact on both overall logistics costs and the service level provided to the customer. Any under-performance in order collection could lead to inadequate service and high operational costs for the warehouse and, eventually, for the supply chain... De Koster (2007) has introduced an efficient configuration of warehouse architecture, storage assignment methods, route methods, order batching, and zoning. The independent analyses of these segments are most prominent and conducted widely, but the combination is almost non-existent. In addition, if congestion happens where several order pickers are available, the time taken to pick up would increase. (Krit Srivilas, 2017).

Hermann Gruenwald (2015) performed that in warehouse operations, order picking prevails over the rest of the operations. Research in the same quarter of 2015 shows that it has a ranking of 37%. So, Order picking is a valuable function. The efficiency of order-picking is split into two major arrays: quality and quantity. This may reflect the mindset of staff, while management may place greater focus on the throughput or products gathered or unloaded. Managers like to insist on delivering the right things at the right time and always want it just in time (JIT), which is an environment with all facets of warehouse operations.

According to Dallari, Marchet, and Melacini (2009), the selection and design of an Order picking system is a very complex task that depends on several factors such as products such as number, size, value, packaging, inventory level, and sales, customer orders and different types of functional areas (e.g. separate areas for fast-moving-product case picking versus slow-moving-product case picking). On the other hand, Makusee Masae (2019) said that picking up pickers can waste more than 50% of your overall order pick-up time traveling around the factory.

From the importance of order picking, several approaches have been developed in the past years to optimize pickers' traveling distance. This paper would rely on the strategies of pick-up-routing.

According to DeKoster (2007) study, order-picking systems are classified based on whether they are operated by humans or automatic machines. Human-use systems are most applicable. On the other hand, De Koster (2004) also points out that picker-to-parts systems are most commonly used, the order picker walks or drives along the aisles to pick it. They are classified into two types: low-level picking and high-level picking. The order picker in a low-level order-picking machine picks ordered items from stock racks or bins (bin-shelving storage) when running around a conveyor belt the aisles of stock. Other order-picking devices make use of high storage racks; order pickers ride to pick positions on the back of a lifting order-pick vehicle truck or crane. The crane would immediately stop ahead of the right location and wait for the recipient to pick the products. This kind of system is referred to as a senior or aboard pick up system. Part-to-pickers systems, like automatic storage and retrieval (AS/RS) systems, are often found in aisle mounting cranes, which take one or more units load (pallet or crate; in the latter case, a device commonly referred to as the small payload) and transport them to the desired position (ie. warehouse). Put systems, also known as order delivery systems are made up of a retrieval and distribution mechanism. A retrieval and allocation mechanism is used in a less widely used order picking method for placed systems. Firstly, the entries must be retrieved, which can be accomplished by selecting each part or by selecting each part. Secondly, items to carry (typically crates) with these the order picker receives pre-selected units and distributes them in accordance with customer orders. Because of this feature, e-commerce warehouses often choose this method to save time packaging. In well-managed schemes, this can result in about 500 picks per order picker hour (for small items). For instance, at the Amazon Germany warehouse or flower auctions) which can result in about 500 picks per order picker hour (for large items).

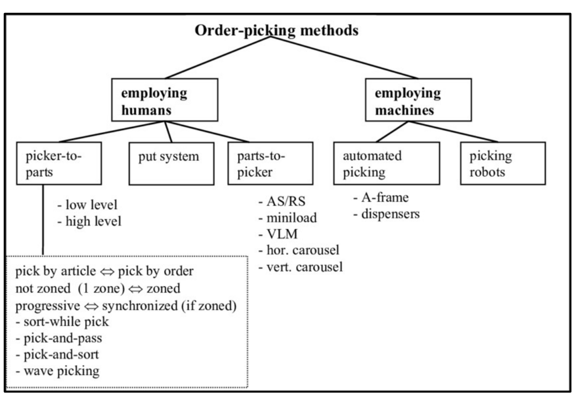


Figure 3. Classified order picking systems

To apply the method appropriate for the study, the author has restricted the article related to low-level, picker-to-parts order picking systems employing humans, as these systems constitute a very large majority of picking systems in warehouses worldwide, especially in the company's current situation.

Order picker routing procedures suggested can be divided into optimal algorithms, heuristics, and metaheuristics.

Chan et al. (2007) proposed new mathematical geometry, with capacity constraints and multiple objectives - a genetic algorithm. The scientific model for the warehouse order optimization problems (W-OPP) is useful. The enhanced GA for the mathematical model is accurate, consistent and can increase the operational efficiency of order collection in warehouses, which also saves resources

Chen et al. (2013) created a routing mechanism to monitor the picker in a narrow-aisle picker scheme. The author has set up a new routing algorithm using the Ant Colony Optimization (ACO) tool for two order pickers called A-TOP. A-function, TOP's which allows the order picker to pass to another block even if the job is not yet done in the current block, starts to play an important role in optimization. When using S-Shape+ and S-Shape, the order picker must visit all the picks in one block before switching to another, without any versatility. In conclusion, in most cases, A-TOP enjoys the fastest overall pick-up period and does well in dealing with congestion. At the same time, this algorithm was doing well in dealing with traffic.

Emir Zunic et al. used the case study Traveling Salesman. The issue of finding the shortest distance can be viewed as a symmetrical TSP, since distances inside the warehouse are equal to the direction we take. The algorithm is dynamic compared with 2-opt and 3-opt heuristics. Using the TPS approach used, the worker's route was shortened by 41.3 percent relative to the previous algorithm. Any more sophisticated algorithms have been developed for TSP, but they were not needed to be introduced, as this would only reduce the path of staff by 1%.

Xiao and Zheng (2009) conducted a series of numerical studies to compare six routing methods for order pickers in single-block warehouses after reviewing other heuristic methods: In a random storage case, the S-shape, return, largest gap, mid-point, composite, and optimum are all possible.

Hall (1993) stated that one of the most basic heuristics for routing order pickers is the S-shape (or traversal) algorithm. When the S-shape algorithm is used, every aisle with at least one pick is fully traversed (except the last visited aisle). Aisles that do not have picks are not joined. The order picker returns to the warehouse from the last aisle visited. This algorithm is effective when entering and leaving data. The largest-gap algorithm outperforms the S-shape algorithm in cases where entering aisles is not time consuming and the density of pick products per aisle is limited. Both aisles except the first and last visited are left on the same side as they were reached in this technique. The order picker returns to the spot with the greatest distance between two neighboring locations to be visited in the aisle.

Dekoster (2007) suggested that, for Return method, where an order picker enters and leaves each aisle from the same end. Only aisles with picks are visited. The midpoint approach splits the warehouse into two parts. Picks in the front half are accessed via the front cross aisle, while those in the back half are accessed via the back cross aisle. The order picker travels to the back half of the store by either the last or first aisle visited. Aisles of picks are either fully traversed or entered and exited at the same end for the mixed (or composite) heuristic.

In this study, the author wants to use S-Shape to build a path for the picker and limit traffic congestion. Moreover, due to the effectiveness of the thesis that can be used in practice, Schneider Company wants to use simple routing policies is considered to be useful especially in systems with many pickers. As a result, this study would use the Mix integer linear programming (MILP) formulation suggested by Öncan and Cağırıcı (2013). The authors introduce a MILP formulation for the order batching problem of S-shape routing and considers the case where the cumulative number of visited aisles is odd. Pickers can join all aisles containing required picking pieces, beginning and ending at the I/O mark. The order pickers may enter an aisle from one end and exit from the other. If the total number of aisles reached is an odd number, the picker will enter and exit from the last aisle before selecting the final pieces. Notably, the picker would not have to walk the whole last aisle in this situation.

3. Data collection and methods

3.1 Conceptual Design Description

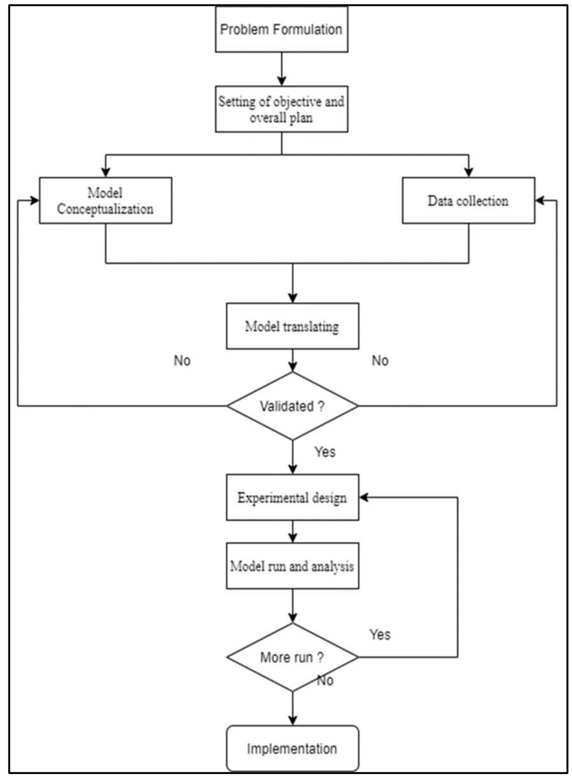


Figure 4. Framework of this paper

Step 1: Problem Formulation: make sure the mathematical model will be built on purpose and doesn't have any flaws. In this study, a mathematical model is simulated to improve warehouse order picking. This first step will form the problem formulation.

Step 2: Setting of objective and overall plan: define the scope and objectives of the case study challenge. The project measures the workload, the number of goods and orders entered and out in the factory. The objectives shall be considered time, expense, and the number of rows to be included.

Step 3: Model Conceptualization: depending on the problem scenario and the goals of the analysis, a first conceptual model can be created. It can be viewed as a flow map or diagram. In this case study, the computational simulation model is the basic method for the warehouse activity of the case study, such as the order collection process.

Step 4: Data collection: working methods, statistical distribution, and critical variables are documented in depth to help the construction of mathematical models in the field.

Step 5: Model translating: based on the contribution of the evaluated data and the mathematical flow map of the model. The simulation model is drawn from the computational model using method modules and data analysis. It was presented with the assistance of IBM Cplex. Through importing data collected in the Excel file to the Input Analyzer, an appropriate distribution for and form of data is collected.

Step 6: Validation of models: validation experiments are used to evaluate the probability of applying model system behavior to the actual system. The statistical model can be tested at a small distance between the actual data and the output data of the model.

Step 7: Experimental design and output analysis: from the performance analysis, it is important to propose changes to the model. Several criteria need to be changed for progress, such as distance travel, pick list, tools, etc.

Step 8: Model runs and documenting: using a statistical approach to interpret the data. The state of critical variables should be assessed in order to find any issues in the framework. Subsequently, propose modifications to the software are updated and run to observe any updates and improvements to the model.

Step 9: Implementation: on the basis of the change, a number of solutions to the problems are presented to the Schneider Electric Company.

3.2 Data Collections

This section will introduce the data collected in the warehouse of Schneider Electric Company for 2.5 years will be used to study this thesis. However, the thesis includes many different processes, so it is inevitable that the data is processed manually so that the output of this process is consistent with the input of the next process. The data types used as inputs for the model are described below:

- The number of orders per day in one month
- The frequency of each category
- The storage required of each category
- Distance from slot to depot
- Time traveled from slot to depot
- The current layout of the warehouse

A sufficient volume of data must be available to construct a mathematical model and test a model. A lack of sufficient evidence is often cited as an explanation for attempting to validate an ineffective model. The author uses 0.05 as the degree of importance for the test in this paper.

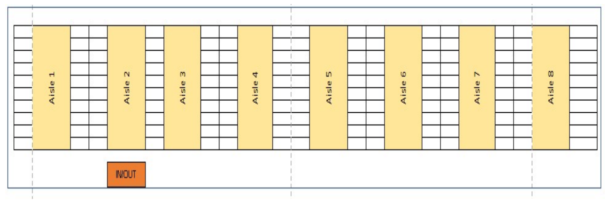


Figure 5. Warehouse layout

3.3. Method

To find out and test current warehouse layout possibilities, there are 2 stages in this study, first of all, using the frequency items to check itemset mining correlate items and align them closely, then construct the MILP formula for the order batch problem with the S-shape routing policy is used to test the optimization of the new item location.

Phrase 1: Frequency itemset mining

Notation:

k-itemset: An itemset having k items.

L_k: Set of large k-itemsets (those with minimum support). Each member of this set has two fields: i) itemset and ii) support count.

C_k: Set of candidate k-itemsets (potentially large itemsets). Each member of this set has two fields: i) itemset and ii) support count.

$\bar{C}_k$: Set of candidate k-itemsets when the TIDs of the generating transactions are kept associated with the candidates.

Model Apriori:

$L_k = \{large\ 1 - itemsets\}$;

for (k =2; $L_k \neq 0$; $k++$) **do begin**

$C_k = apriori - gen\ (L_{k-1})$;// New candidates

forall transactions $t \in \boldsymbol{D}$ **do begin**

$C_t = subset\ (C_k, t)$; // Candidates contained in t

forall candidates $c \in C_t$ *do*

c.count++;

end

$L_k = \{c \in C_k\}\ c.count \geq minsup$

end

Answer = $\cup_k l_k$

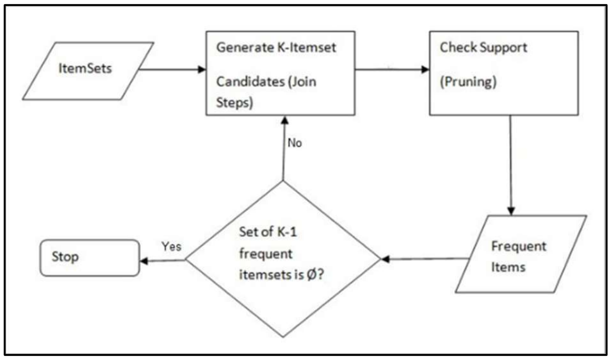


Figure 6. FlowChart of Apriori

Step 1: Each object is treated as a 1-itemsets nominee in the first iteration of the algorithm. Each item's occurrences will be counted by the algorithm.

Step 2: Allow for some minimum assistance, min-sup. The set of 1 – itemsets whose frequency meets the min sup condition is decided. Only candidates with a score greater than or equal to min sup are considered for the next iteration, while the others are pruned.

Step 3: Following that, 2-itemset frequent items with min-sup are found. The 2-itemset is formed in the join phase by creating a group of 2 by combining items with itself.

Step 4: The min-sup threshold value is used to prune the 2-itemset applicants. The table will now only have two –itemsets with min-sup.

Step 5: The next iteration would use a join and prune move to create three –itemsets. This iteration will obey the antimonotone property, where the subsets of 3-itemsets, that is, the 2 –itemset subsets of each category, fall in min-sup. If all 2-itemset subsets are frequent, then the superset is frequent; otherwise, it is pruned.

Step 6: The final move would be to join 3-itemset with itself and prune if its subset does not satisfy the min sup criterion. When the most regular itemset is reached, the algorithm is terminated.

Phraise 2: MILP formulations for the order batching problem

Parameters:

- i = set of orders for i=1,2,3,..n;
- k = aisles for k=1,2,..,K;
- l = aisles front of in/out point for l=1,..K;
- j = number of batch processing for j=1,2,3,..J;
- L = horizontal length of the aisles;
- w = the width between aisles ;
- d_{ik} = maximum vertical distance from front of aisle k to the location of item requested in order I; d_{ik} = 0 when no order i has picked in aisle k.
- m_i = number of items in order i;

Decision Variables:

- x_{ij} = *1,* if and only if order i is assigned belong to batch j; otherwise,0.
- x_{jj} = *1,* if and only if order j is assigned belong to batch j; otherwise,0.
- $y_{jk} = 1$, if the picker I serving batch j to aisle k; otherwise,0.
- $c_j = 1$, if the picker I serving batch j to aisle k; otherwise,0.
- $p_{jk} = 1$, if the aisle k is the furthest aisle that the pickers visited ; otherwise,0.
- $v_j = 1$, total number of time visited by picker, two-way travelsals.
- $h_j^R =$ one way horizontal from the aisle front of I/O point to the right.
- $h_j^L =$ one way horizontal from the aisle front of I/O point to the left.
- $u_{jk} =$ horizontal one distance travel in the furthest aisle k by picker serving batch j,if total vistied aisles is odd

Objective Function:

(1) $Min\ Z = 2\sum_{j=1}^{J}\sum_{k=1}^{K}u_{jk} + \sum_{j=1}^{n}(h_j^R + h_j^L) + 2L\sum_{j=1}^{n}(v_j - c_j)$;

(2) $\sum_{j=1}^{J}x_{ij} = 1$ *for i=1,2,..n;*

(3) $x_{ij} \le x_{jj}$ *for i, j=1,2,..n;*

(4) $\sum_{i=1}^{n}m_i x_{ij} \le Q$ *for j=1,2,..n;*

(5) $y_{jk} \le \sum_{j=1}^{n}d_{ik}x_{ij} \le My_{jk}$ *for j=1,2,3,..n; k=1,2,..K;*

(6) $(k-l)wi_{jk} \le h_j^R$ *for j=1,2,..n; k=l+1,..K;*

(7) $(l-k)wi_{jk} \le h_j^L$ *for j=1,2,..n; k=1,..l-1;*

(8) $\sum_{k=1}^{K} y_{jk} + c_j = 2v_j$ *for j=1,2,..n;*

(9) $d_{ik}x_{ij} \le u_{jk} + M(1 - p_{jk}) + M(1 - c_j)$

for i, j=1,2,3,..n; k=1,2,..K;

(10) $y_{jk} - \sum_{t=k+1}^{k} y_{jl} \le p_{jk} \le y_{jk}$

for j=1,2,3,..n; k=1,2,..K;

(11) $v_j: integer\ for\ j = 1,2,..,J$

(12) $x_{ij} \in \{0; 1\}\ for\ i = 1,2,..,n;\ j = 1,2,..,J$

(13) $p_{jk} \in \{0; 1\}\ for\ j = 1,2,..,J;\ k = 1,2,..,K$

(14) $y_{jk} \in \{0; 1\}\ for\ j = 1,2,..,J;\ k = 1,2,..,K$

(15) $c_j \in \{0; 1\}\ for\ j = 1,2,..,J$

(16) $h_j^R \ge 0\ for\ j = 1,2,..,J$

(17) $h_j^L \ge 0\ for\ j = 1,2,..,J$

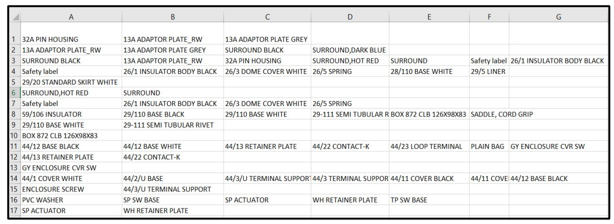
(18) $u_{jk} \ge 0\ for\ j = 1,2,..,J;\ k = 1,2,...,K$

In the goal of this work (1) the first illustrates the distance in the furthest right walkway when a picker absolutely visits an odd number of paths. Next is for the even distance of pickers traveling. The third term clarifies for the absolute distance that went by the pickers which shows two-way whole crossings from the front passageway to the back path and from the back walkway to the front passageway. Constraints (1) Each unique batch will be identified through an order. Turn to the next constraints (2) present for the orders cannot be delegated to the batch j represented by order j. Constraints (3) ensure that the products in all orders allocated to batch j are beyond the picker's capability. Furthermore, when there is at least one thing in aisle k that belongs to order I batch j, including order I must join the aisle k (constraint 4). Constraints (5) present that when there exists somewhere around one thing which has a place with request i and which is situated in walkway k, then, at that point the picker serving cluster j which incorporates request I, should go into passageway k. In this function, M in Big M is a sufficiently large number. By default, if no command from batch J goes to k, the y_{jk} value is also 0. For constraints 6 and 7, it is used to consider horizontal distance when j starts from the I/O point along the path forward. left or right (this point is optimal when the I/O position is in the middle or not in the 1st aisle at the warehouse). Constraints (8) ensure where the number of aisles traversed by a picker serving batch j is odd, set $c_j = 1$ to. Constraints (9) If the total number of aisles visited is odd, calculate the one-way vertical distance traveled in the furthest aisle visited by batch j. Constraints (10) Declare that p_{jk} should be equal to 1 while aisle k visited by the picker serving batch j is the picker's rightmost aisle visited. At last, the domain definition of decision variables has shown from constraints 11 to 18.

4. Result

4.1 Apriori algorithm result

Data input: To implement the Apriori algorithm, the input data will be modified through excel, the final result is presented in excel. This section will explain how to input data and output data. Input data: Materials are listed by order, excluding the quantity of each item.



	A	B	C	D	E	F	G
1	32A PIN HOUSING	13A ADAPTOR PLATE_RW	13A ADAPTOR PLATE GREY				
2	13A ADAPTOR PLATE_RW	13A ADAPTOR PLATE GREY	SURROUND BLACK	SURROUND,DARK BLUE			
3	SURROUND BLACK	13A ADAPTOR PLATE_RW	32A PIN HOUSING	SURROUND,HOT RED	SURROUND	Safety label	26/1 INSULATOR BODY BLACK
4	Safety label	26/1 INSULATOR BODY BLACK	26/3 DOME COVER WHITE	26/5 SPRING	28/110 BASE WHITE	29/5 LINER	
5	29/20 STANDARD SKIRT WHITE						
6	SURROUND,HOT RED	SURROUND					
7	Safety label	26/1 INSULATOR BODY BLACK	26/3 DOME COVER WHITE	26/5 SPRING			
8	S9/106 INSULATOR	29/110 BASE BLACK	29/110 BASE WHITE	29-111 SEMI TUBULAR R	BOX 872 CLB 126X98X83	SADDLE, CORD GRIP	
9	29/110 BASE WHITE	29-111 SEMI TUBULAR RIVET					
10	BOX 872 CLB 126X98X83						
11	44/12 BASE BLACK	44/12 BASE WHITE	44/13 RETAINER PLATE	44/22 CONTACT-K	44/23 LOOP TERMINAL	PLAIN BAG	GY ENCLOSURE CVR SW
12	44/13 RETAINER PLATE	44/22 CONTACT-K					
13	GY ENCLOSURE CVR SW						
14	44/1 COVER WHITE	44/2/U BASE	44/3/U TERMINAL SUPPORT	44/3 TERMINAL SUPPOR	44/11 COVER BLACK	44/11 COVE	44/12 BASE BLACK
15	ENCLOSURE SCREW	44/3/U TERMINAL SUPPORT					
16	PVC WASHER	SP SW BASE	SP ACTUATOR	WH RETAINER PLATE	TP SW BASE		
17	SP ACTUATOR	WH RETAINER PLATE					

Figure 7. The order item lists to relayout

In case : min_support = 0.1, min_confidence = 0.5, min_lift = 2, min_length= 2.

Result below consider by testing 20 SKU:

Items group		Statistics
Item 2	Item 4	0.43
Item 7	Item 3	0.23
Item 2	Item 10	0.21
Item 5	Item 6	0.12
Item 18	Item 22	0.12

Table 1. Result from Apriori algortihm.

The results show that items with high frequency will often be ordered together in order lists. Using the Apriori algorithm for grouping will reduce travel time as well as distance. Although, when they had to go twice like the original, now the distance is only 1/2.

4.2 MILP formulations for the order batching problem results

Number of orders	10
Number of batch	10
Number of aisles	8
In/out point	2
Vertical length of aisle (m)	11
Width between aisle (m)	3
Capacity	40

Table 2. Data input in MILP formulations

Batch	Before	After	Order
1	102	92	1,5
2	68	57	2,6,10
3	80	54	3,7,8
4	22	15	4,9
Total distance	272	218	

Table 3. Result from Cplex by using MILP

Result: Optimal solution for 10 orders: 218 m

After performing all the SKUs by itemsets checking, the total travel distance for the new route is compared with the current layout at the Schneider warehouse. We can clearly see a significant reduction compared to the original design position, so this arrangement achieves the original purpose. In fact, Schneider's original move was Picker dependent so there was no standard, they just looked at the map and moved to collect all the items. Therefore, the difference obtained is quite large, and the potential for application. In addition, the products that are picked up have been located near the In/Out point, so it is convenient to move and reduce the distance for people to pick up items in order picking lists. Therefore, it can be said that this is a feasible approach that can be applied to the current inventory and meets the purpose set out for this paper.

The reduction of travel is optimized during this period, which will not only help workers reduce the amount of travel, but also help the company adjust the number of employees appropriately. Outlining the clear path is saved on the system, will trace back the data to move. This information will help them better manage both employees and health.

Today, without applying the above algorithm, workers will move at will until they get all the materials they need, obviously, it will be very difficult to trace the distance and not trace in cases. emergency. Hopefully, with the results as well as the contributions from the above article, the warehouse management will be improved as well as reduce the cost loss due to the bottleneck stagnation occurring at the Schneider warehouse.

5. Conclusion & Recommendation

5.1. Summary of results

As the result, the all out distance went for each parcel is determined for both old orders, what is more, in the new format, we can see that the absolute distance expected to choose all requests of the last design is essentially more modest than the first format by in excess of 60 meters. Besides, the complete distance of each clump after revamp is a lot more modest when contrasted with the first, which implies that pickers can lessen the moving responsibility. From the above strategy, the outcomes are positive for the methodology at the stockroom, in this way limiting the measure of execution misfortune caused as the first issue proposed. Eminently, if the all out distance went in the new format is as yet greater than the first design, so adjusting everything things until the most ideal arrangement is reached.

5.2. Recommendations

For the proposals mentioned in the article, it is considered as a new vision for the development of the Schneider factory in Vietnam. The development in the direction of 5 Why goes from core values but has a positive impact on Internal Defect Costs for this factory.

The article is used real data and has a Risk assessment assessed by Warehouse Manager and Method Maintenance office in Schneider Electric VietNam to ensure the performance after completion. However, the process of changing the relayout warehouse will never be easy, but the value brought from the results above can draw a great development in this department.

In the supply chain network, a stockroom can be a base for item combination to downsize transportation costs, accomplishing economies of scale in assembling or buying or providing esteem added cycles and cut crude material stand by times. In this way, to be powerful. Unite efficiency and diminish the working expenses of a distribution center, sought after to designate assets in the stockroom productively. In addition, perhaps the main things. The factor that decides the productivity of a stockroom is the assurance of the right stockpiling place area for a large number of things in a distribution center. In a compelling picking strategy, there are will be expanded exactness and decrease work cost if work is recruited. More modest framework care saves work on the distribution center floor; it is likewise extensible capacity to give better client assistance by lessening process durations. Besides, a get strategy won't just prompt a decrease in stock, yet additionally an increment away limit.

The reduction of travel is optimized during this period, which will not only help workers reduce the amount of travel, but also help the company adjust the number of

Reference

Gruenwald, H. (2015). Warehouse Operations from Theory to Industry Best Practices in Thailand. *TNI Journal of Business Administration and Languages*, *3*(2), 63-65.

Ramaa, A., Subramanya, K. N., & Rangaswamy, T. M. (2012). Impact of warehouse management system in a supply chain. *International Journal of Computer Applications*, *54*(1).

Zunic, E., Besirevic, A., Skrobo, R., Hasic, H., Hodzic, K., & Djedovic, A. (2017, October). Design of optimization system for warehouse order picking in the real environment. In *2017 XXVI International Conference on Information, Communication and Automation Technologies (ICAT)* (pp. 1-6). IEEE.

Sivakumar, V., & Ruthramathi, R. (2019). Challenges and features of warehousing operations with respect to logistics warehouse companies in chennai. *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 3304-3308.

Gu, J., Goetschalckx, M., & McGinnis, L. F. (2007). Research on warehouse operation: A comprehensive review. *European journal of operational research*, *177*(1), 1-21.

Tompkins, J. A. (1994). *The distribution management handbook*. McGraw-Hill.

De Koster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European journal of operational research*, *182*(2), 481-501.

Anouze, A. L. M., & Bou-Hamad, I. (2019). Data envelopment analysis and data mining to efficiency estimation and evaluation. *International Journal of Islamic and Middle Eastern Finance and Management*.

Mueller, A. (1998). *Fast sequential and parallel algorithms for association rule mining: A comparison*.

Bhandari, A., Gupta, A., & Das, D. (2015). Improvised apriori algorithm using frequent pattern tree for real time applications in data mining. *Procedia Computer Science*, *46*, 644-651.

Srivilas, K., & Cherntanomwong, P. (2017, March). Routing algorithm in warehouse with congestion consideration using an ACO with VLC support. In *2017 International Electrical Engineering Congress (iEECON)* (pp. 1-4). IEEE.

Masae, M., Glock, C. H., & Grosse, E. H. (2020). Order picker routing in warehouses: A systematic literature review. *International Journal of Production Economics*, *224*, 107564.

Dallari, F., Marchet, G., & Melacini, M. (2009). Design of order picking system. *The international journal of advanced manufacturing technology*, *42*(1-2), 1-12.

CHANG, F. L., LIU, Z. X., Zheng, X. I. N., & LIU, D. D. (2007). Research on order picking optimization problem of automated warehouse. *Systems Engineering-Theory & Practice*, *27*(2), 139-143.

Chen, F., Wang, H., Qi, C., & Xie, Y. (2013). An ant colony optimization routing algorithm for two order pickers with congestion consideration. *Computers & Industrial Engineering*, *66*(1), 77-85.

Dekker, R., De Koster, M. B. M., Roodbergen, K. J., & Van Kalleveen, H. (2004). Improving order-picking response time at Ankor's warehouse. *Interfaces*, *34*(4), 303-313.

Xiao, J., & Zheng, L. (2010). A correlated storage location assignment problem in a single-block-multi-aisles warehouse considering BOM information. *International Journal of Production Research*, *48*(5), 1321-1338.

Hall, R. W. (1993). Distance approximations for routing manual pickers in a warehouse. *IIE transactions*, *25*(4), 76-87.

Öncan, T., & Cağırıcı, M. (2013). MILP formulations for the order batching problem in low-level picker-to-part warehouse systems. *IFAC Proceedings Volumes*, *46*(9), 471-476.

DESIGN SUPPLY CHAIN NETWORK FOR REVERSE LOGISTICS SCANCOM CASE STUDY

Vo Hong Thanh Nhi[1], Anh Duong[2]

[1,2]Department of IEM, International University, Thu Duc, HCMC, Vietnam; Email:

[1]thanhnhi0406@gmail.com, [2]dvnanh@hcmiu.edu.vn

Abstract. Reverse logistics has become a key trend to balance economic development with environmental protection and it is a must to consider where to build the maintenance centers where the product should be sent and the necessary number of them. Identifying optimum location for facility is one of the major challenges in logistics network. In this study, the facility location for maintenance center is proposed to help improve the reverse logistics networks for Scancom. The potential location is defined by the Center of Gravity algorithms, hierarchical clustering algorithm and a Mixed-integer linear programming model to help optimizing the systems.

Keyword: reverse logistics, facility location, hierachical clustering algorithm, COG, MILP

1. Introduction

1.1. Background

1.1.1. Definition of reverse logistics

Reverse logistics has become a key trend to balance economic development with environmental protection (Rogers et al., 2012). Reverse logistics is defined as "The process of planning, implementing, and controlling the efficient, cost-effective flow of raw materials, in-process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal" (Rogers and Timbben-Lembke, 1999).

Thus, reverse logistics includes all activities as forward logistics. However, they operate in reverse cycle. Therefore, reverse logistics is the process of effectively planning, implementing, and controlling the flow of goods and related services and information from the point of consumption back to the origin. The purpose of recovering any residual value of goods or for reasonable disposal.

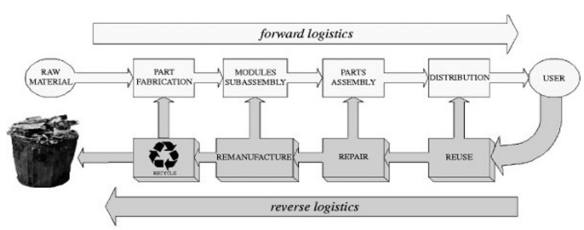


Figure 1: Comparison between forward logistics and reverse logistics

Reverse logistics is formed based on various reasons such as: recalling unsold products for improvement, recalling reusable packaging, recalling defective products, recalling products with partially dismantable and reusable, ...

There are 4 main steps in Reverse Logistics, including:

- Aggregation: The activity of recalling unsold products, defective products, packaging and then transporting them to the point of recall.

- Inspection: At the point of recall, recalled goods will be inspected in terms of quality, selected, and classified according to the criteria. This inspection plays an important role and affects how to do the next step.

- Handling: With the recovered goods, the enterprise has many ways to handle which are to reuse directly or resell, product recovery: repair the defective product, re-manufacture, seperate it to get spare parts, etc. and an important step is to dispose of waste if it is no longer usable (in a way that minimizes the impact on the environment)

- Distributing restored product: ogistics will take place normally with storage, transportation, and sales activities in this step.

1.1.2. Benefit of Reverse Logistics:

In the supply chain, people focus on logistics's role as it determines product quality, competitive prices, reasonable product life cycle, on-time delivery, and failure rate. Low failures are crucial and will continue to play a more important role in the future.

But at present, the recall of goods is an obvious problem for manufacturers, exclusive distribution intermediaries, wholesalers, traditional retailers, and online retailers as well as Logistics service providers. Reverse logistics will be a way to reduce costs, increase revenue and improve customer service, thereby helping the company gain a competitive advantage in the market.

1.2. Problem Statement – The Need of Study:

In this case study, there is an average of 3% of defective products a year out of the total number of furniture products such as tables, chairs, etc. produced by the branches of Scancom companies around the world are returned from the customers. Moreover, statistically, this 3% has caused a loss of profits for the company of up to 60 billion $. With the aim of reducing consumption and develop customer service and satisfaction also in line with the environmental concerns and the forward logistics support, this paper proposes the company to come up with a supply chain design for the Reverse logistics system for Scancom worldwide. This reverse logistics system is created with the purpose of collecting defective and damaged items and sort them so that they can be recovered or properly disposed of.

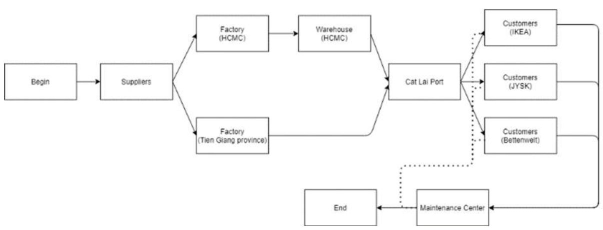


Figure 2: Flowchart for the process of how reverse logistics of Scancom works.

The reverse logistics network discussed in this paper is a multistage logistics network including customers, maintenance center, recycling center, and disposal center. More specifically, as figure 2 has illustrated the process of how the system works, the returned product will be collected from the customers into the maintenance center and after the inspection process; these items will be divided into recoverable products and scrapped products. Products that can be recovered will enter the reverse flow into the recovery center and scrapped products will be disposed or resell to second customer market. The recovered products will be sent back to the customers. So, there is a must to consider where to build these colection centers where the product should be sent and the necessary number of them.

1.3. Objectives of Study:

This paper aims to design a reverse logistics system by applying hierarchical clustering method and a mixed integer linear programming to help improving the satisfaction of customers, minimizing total costs in the process of handling defective products to provide economic and environmental benefits.

First, reverse logistics supports customer services by getting defective products back into the reverse logistics line. Moreovers, the reverse logistics help to utilize 3% of defective returned products to increase trust from customers.

Next, reverse logistics protects the environment by quickly recalling defective products and sorting them for recovered or treating them as waste safely and efficiently. Additionally, returning the used products that no longer in use from customers back to the reverse flow of logistics also helps to save costs for raw materials and reduce environmental pollution, especially deforestation.

Finally, the project is aimed at minimizing the total cost so that reverse logistics systems can be available.

2. Literature Review

Jimenez et al. (2019) proposed a publication to spot good practices and trends within the plastic products manufacturing industry by applying the AHP method to conclude that Leadership factor is the most important factor.

Most of the publications about logistics network design considers various facility models based on the mixed integer linear programming. Likewise, the literature review will introduce a few articles that provide some models for reverse logistics.

According to the reverse logistics literature, the uncertainty is the most vital factor to controls. Jayaraman and Ross (2003) proposed a MILP model which is a formal closed-loop logistics system and a single time-period model, REVLOG for reverse logistics network. This model is designed to deal with inventory explosion or significant inventory backlogs due to the unstable demand for remanufactured products. The objective of the proposed model is to attenuatte the total costs. The bonus point of this model is the ability to identify the parameters that influence system design.

Lieckens and Vandaele (2007) transformed the traditional model into a mixed integer nonlinear program model (MINLP) during a single level, single product network design problem that covers a single period. With the aim of incorporating a product's cycle time and congestion and inventory holding costs, as well as to deal with the higher degree of uncertainty and congestion, typical characteristics of these networks.

Rogers et al. (2012) applied modeling methodologies to manage reverse logistics problems and issues. The paper considered several key reverse logistics problems.

Pishvaee et al. (2010). proposed a MILP model to minimize the transportation and fixed opening costs of a multistage logistics network using a simulated annealing (SA) algorithm and designed an integrated logistics network design model to avoid the sub-optimality caused by a separate, sequential design of forward and reverse logistics networks. Their bi-objectivemixed integer programming model was designed to minimize the total costs and to maximize the responsiveness of a logistics network (Pishvaee, 2010)

In addition to the above methods, a clustering analysis is also a method widely used in data processing and is applied in many different specialties. The purpose of cluster analysis (also known as classification) is to construct groups (or classes or clusters) while ensuring the following property: within a group the observations must be as similar as possible, while observations belonging to different groups must be as different as possible. In facility location – allocation, clustering analysis play an important role in finding the appropriate location. There are two mains widely used types of clustering analysis which are k-means clustering and hierachical clustering. Sheu and Lin (2012) proposed a novel hierarchical network planning model for global logistics (GLs) network configurations with the aim of determining the corresponding locations, number and scope of service areas and facilities in the proposed GLs network.

Based on the economical concerns, Kanna et al. (2012) proposed a MILP model which is the extension version of simulated annealing (SA) algorithm from Pishvaee et al. (2010) for a carbon footprint based reverse logistics design. The proposed model aims at minimizing carbon footprint emissions and the costs in reverse logistics combining the location or transportation decision.

Kuşakci et al. (2019) investigated the fuzzy mix-integer location-allocation model for reverse logistic network of end-of-life vehicles by applying them to the case study in Turkey. Sensitivity analysis is also applied to the number of required processing facilities and the total profit by changing the amount of returned products. The objective of the paper is to determine the location of end-of-life vehicles treatment facilities and the material flows between the clusters of the network for different types of subcomponents and materials of the vehicles.

Azizi et al. (2020) designed a two-stage stochastic programming model for multi-echelon multi-period reverse logistics including lot-sizing (including backorder and shortage) and outsourcing while considering return and demand uncertainty as a stochastics factor. The optimal numbers of sorting center and warehouses, optimal lot sizes, transportation plan that minimize the expected total system cost determination is the objective of the paper.

In line with the COVID-19 pandemic, Kargar et al. (2020) developed a linear programming model which created a balance between three objectives for the reverse logistics networks designed for medical waste management in the pandemic outbreak of COVID-19. In this research, minimizing the total costs, the risk associated with the transportation and treatment of infectious medical waste and the maximum amount of uncollected waste in medical waste generation centers are the objectives of the proposed model.

Liao (2018) applied a generic mixed integer nonlinear programming model which is a hybrid genetic algorithm (GA) for multi-echelon reverse logistics design. Maximizing the total profit with consideration of multi-product and multi-module returnable products and a variety of recycling channels for the returned products is the objective of the paper.

Parchami Afra and Behnamian (2021) modeled the multi-product production routing problem with opening cost and environmental consideration as mixed-integer linear programming. NP-hardness and Lagrangian Relaxation algorithms are applied in this paper to gain the objective of the balancing of minimizing the cost and carbon emission.

Trochu and Ouhimmou (2018) proposed a mixed-integer linear programming model (MILP) to analyse the direct impact of different key uncertain parameters on RL network design decisions under the waste management policy. The objective is minimising the total cost of the wood recycling process collected from renovation and demolition sites.

It can be concluded that there are many problems solving facility location problems, of which the three articles giving the most efficient and appropriate model in this case study belong to Vahid Azizi et al, and Pishvaee et al, together with mathematical models respectively including a Two stage Stochastic, and a Simulated Annealing algorithm and the K-means clustering algorithm. Among three model, clustering analysis works the best for the problem considered in this paper.

While there are two main types of clustering algorithm which are k-means clustering and hierarchical clustering, a comparison table is conducted as below:

Table 1. Comparision between k-means clustering and hierarchical clustering

k-means clustering	Hierarchical clustering
The time complexity of K Means is linear	The time complexity of hierarchical clustering is quadratic
Must predefine the quantity of clusters	Determine the number of clusters should be divided
Big data with sizable amount of variables	Small data
Centroid based, partition-based	Hierarchical, Agglomerative
Use the Elbow method using WCSS to seek out the optimal number of clusters	Use Dendrogram to find the optimal number of clusters
Does not have diractional approach because the only centroid is taken into account to create cluster	Top – down, bottom – up

Due to the geodata of the issues, this research paper prefers hierarchical clustering.

3. Modeling:

In this research paper, hierarchical clustering is applied to the case study of reverse logistics in Scancom so as to determine the optimal number of facility location for maintenance center and a MILP mathematical model is developed with the aim of minimizing the total cost icluding the fixed opening cost of maintenance center and transportation costs between the facilities for the reverse logistics systems. To specify the scope for the study, some assumptions and simplications for the proposed model formulation are postulated as follows:

- All the returned products from the customers must be collected.
- Customer locations are fixed and predefined.
- Number, locations, and capacities of maintenance center are known in advance.

The following flowchart are the steps for the solution development in this study.

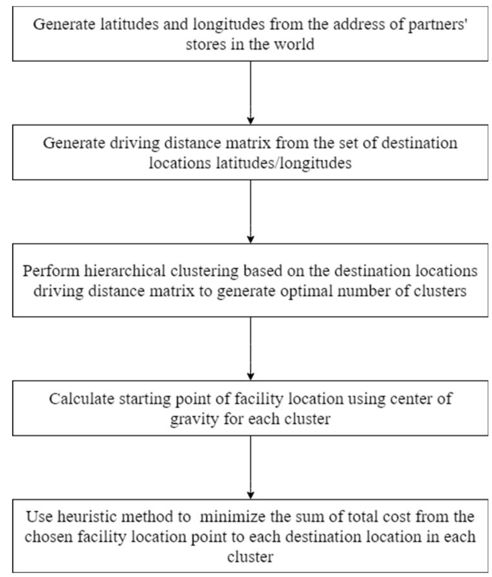


Figure 3: Solution development steps flowchart

- Step 1: Generate latitudes and longitudes from the address of partners' stores in the world.

Country	Latitude	Longitude
Australia	25.000	135.000
Austria	46.527	9.600
Bahrain	25.943	50.601
Belgium	50.501	4.476
Bosnia-Herzegovina	43.915	17.679
Bulgaria	42.725	25.483
Canada	62.227	(105.381)
China	35.859	104.136
Croatia	44.474	16.469
Cyprus	35.095	33.203
Czech Republic	49.195	16.607
Denmark	55.676	12.568
Dominican Republic	18.483	(69.930)
Egypt	30.033	31.233
Estonia	58.597	24.987
Finland	61.924	25.748
France	48.865	2.349

Figure 4: Generate latitudes and longitudes by Excel VBA.

- Step 2: Generate driving distance matrix from the set of destination locations latitudes/longitudes.

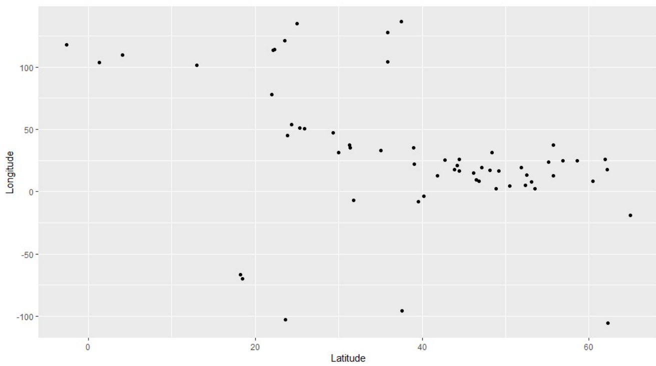


Figure 5. All store distribution

- Step 3: Perform hierarchical clustering based on the destination locations driving distance matrix to generate optimal number of clusters. The result of applying this algorithm is a dendrogram and from this dendrogram together with several tests, the number of clusters and clusters are shown.
- Step 4: Calculate starting point of facility location using center of gravity for each cluster. Using latitude and longitude coordinates together with the number of stores coordinated in each country to calculate the initial facility location centers for each cluster. And by the means of using Google Map, the initial facility location is identified from the calculated x-coordinates and y-coordinates from the COG algorithm. The COG algorithm is as follows:

$$C_x = \frac{\sum X_i \times W_i}{\sum W_i}$$

$$C_y = \frac{\sum Y_i \times W_i}{\sum W_i}$$

- Step 5: Use heuristic method to minimize the sum of total cost from the chosen facility location point to each destination location in each cluster.

3.1. Hierarchical Clustering algorithm:

Hierarchical clustering methods start with all patterns as one cluster and successively perform splitting or merging until a stopping criterion is met. This ends up in an exceedingly tree of clusters called dendograms. The dendogram could also be cut at different levels to yield desired clusters. Well-known hierarchical clustering algorithms include balanced iterative reducing and clustering using hierarchies (BIRCH), clustering using interconnectivity (Chameleon), clustering using representatives (CURE), and robust clustering using links (ROCK (Huang, 2017).

Hierarchical clustering is as simple as K-means, but rather than there being a collection of number of clusters, the quantity changes in every iteration. If the quantity increases, divisive clustering is discussed as: all data instances start in one cluster, and splits are performed in each iteration, leading to a hierarchy of clusters. On the other hand, agglomerative clustering could be a bottom-up approach: each instance is a cluster at the beginning, and clusters are merged in every iteration. With use of either method, the hierarchy will have $N - 1$ levels.

This way, hierarchical clustering does not provide a single clustering of the information but provides clustering of $N - 1$ of them. It is up to the user to determine which one fits the aim. Statistical heuristics are sometimes employed to aid the decision.

The arrangement after training, the hierarchy of clusters, is usually plotted as a dendrogram. Nodes within the dendrogram represent clusters. The length of a grip between a cluster and its split is proportional to the dissimilarity between the split clusters. The recognition of hierarchical clustering is expounded to the dendrograms: these figures provide an easy-to-interpret view of the clustering structure (Biamonte et al., 2017).

3.2. Mathematical model:

Set

- J set of demand
- K set of demand market location.
- L set of potential maintenance center

Parameter

- A_j production cost of product j
- BH_j transportation cost of product j per km from Factory (HCMC) to warehouse
- BT_j transportation cost of product j per km from Factory (TG) to warehouse
- CW_j transportation cost of product j per km from warehouse to Cat Lai port

- CT_j transportation cost of product j per km from Factory (TG) to Cat Lai port
- DI_j transportation cost of product j per km from Cat Lai port to IKEA
- DJ_j transportation cost of product j per km from Cat Lai port to JYSK
- DB_j transportation cost of product j per km from Cat Lai port to Bettenwelt
- EI_j transportation cost of product j per km from IKEA to maintenance centers
- EJ_j transportation cost of product j per km from JYSK to maintenance centers
- EB_j transportation cost of product j per km from Bettenwelt to maintenance centers
- F_l fixed cost for opening maintenance center l
- G_j cost saving of product j (due to product recovery)
- H_j disposal cost of product j
- P_{lj} capacity of maintenance center l
- d_{Ij} demand of IKEA for product j
- d_{Jj} demand of JYSK for product j
- d_{Bj} demand of Bettenwelt for product j
- r_{Ij} return of IKEA for product j
- r_{Jj} return of JYSK for product j
- r_{Bj} return of Bettenwelt for product j
- t_{HWj} distance from from Factory (HCMC) to warehouse
- t_{TWj} distance from from Factory (TG) to warehouse
- t_{WPj} distance from from warehouse to Cat Lai port
- t_{TPj} distance from from Factory (TG) to Cat Lai port
- t_{PIj} distance from from Cat Lai port to IKEA
- t_{PJj} distance from from Cat Lai port to JYSK
- t_{PBj} distance from from Cat Lai port to Bettenwelt
- t_{ICj} distance from from IKEA to maintenance centers
- t_{JCj} distance from from JYSK to maintenance centers
- t_{BCj} distance from from Bettenwelt to maintenance centers

Variables

- Q_{HWj} quantity of product j delivered ftom Factory (HCMC) to warehouse
- Q_{TPj} quantity of product j delivered ftom from Factory (TG) to Cat Lai Port
- Q_{WPj} quantity of product j delivered ftom from warehouse to Cat Lai Port
- Q_{PIj} quantity of product j delivered ftom from Cat Lai Port to IKEA
- Q_{PJj} quantity of product j delivered ftom from Cat Lai Port to JYSK
- Q_{PBj} quantity of product j delivered ftom from Cat Lai Port to Bettenwelt
- Q_{ICj} quantity of product j delivered ftom market demand IKEA to maintenance center l.
- Q_{JCj} quantity of product j delivered ftom market demand JYSK to maintenance center l.
- Q_{BCj} quantity of product j delivered ftom market demand Bettenwelt to maintenance center l.
- QI_{lrj} quantity of returned product j delivered ftom maintenance center l back to IKEA.
- QJ_{lrj} quantity of returned product j delivered ftom maintenance center l back to JYSK.
- QB_{lrj} quantity of returned product j delivered ftom maintenance center l back to Bettenwelt.
- $Z_l = \begin{cases} 1 \text{ if collection center } l \text{ is opened} \\ 0, \text{otherwise} \end{cases}$

Objective function

$$
\begin{aligned}
Min\ Z = &\sum_{j\in J} F_l \times Z_l + \sum_{j\in J} (A_j + BH_j \times t_{HWj}) \times Q_{HWj} \\
&+ \sum_{j\in J} (A_j + CT_j \times t_{TPj}) \times Q_{TPj} \\
&+ \sum_{j\in J} (A_j + CW_j \times t_{WPj}) \times Q_{WPj} \\
&+ \sum_{j\in J} (A_j + DI_j \times t_{PIj}) \times Q_{PIj} \\
&+ \sum_{j\in J} (A_j + DJ_j \times t_{PJj}) \times Q_{PJj} \\
&+ \sum_{j\in J} (A_j + DB_j \times t_{PBj}) \times Q_{PBj} \\
&+ \sum_{l\in L} \sum_{j\in J} EI_j \times t_{IC} \times Q_{ICj} \\
&+ \sum_{l\in L} \sum_{j\in J} EJ_j \times t_{JC} \times Q_{JCj} \\
&+ \sum_{l\in L} \sum_{j\in J} EB_j \times t_{BC} \times Q_{BCj} \\
&+ \sum_{l\in L} \sum_{j\in J} (-GI_j + EI_j \times t_{ICj}) \times QI_{lrj} \\
&+ \sum_{l\in L} \sum_{j\in J} (-GJ_j + EJ_j \times t_{JCj}) \times QJ_{lrj} \\
&+ \sum_{l\in L} \sum_{j\in J} (-GB_j + EB_j \times t_{BCj}) \times QB_{lrj}
\end{aligned}
$$

Constraints

$$Q_{HW} + Q_{TW} + Q_{WP} = Q_{PIj} + Q_{PJj} + Q_{PBj} \qquad (1)$$

$$Q_{PIj} + Q_{PJj} + Q_{PBj} \geq d_{Ij} + d_{Jj} + d_{Bj} \quad (2)$$

$$QI_{lrj} + QJ_{lrj} + QB_{lrj} \leq Q_{lj} \quad (3)$$

$$QI_{lrj} + QJ_{lrj} + QB_{lrj} \leq Q_{PIj} + Q_{PJj} + Q_{PBj} \quad (4)$$

$$Q_{ICj} + Q_{JCj} + Q_{BCj} \leq Z_l \times Q_{lj} \quad (5)$$

$$\sum_{l \in L} Q_{ICj} = r_{Ij} \quad (6)$$

$$\sum_{l \in L} Q_{JCj} = r_{Jj} \quad (7)$$

$$\sum_{l \in L} Q_{BCj} = r_{Bj} \quad (8)$$

$$Z_l \in \{0; 1\} \quad (9)$$

Q_{HWj}, Q_{TWj}, Q_{Icj}, Q_{JCj}, Q_{JCj}, Q_{lj}, QI_{lrj}, QJ_{lrj}, QJ_{lrj}, Q_{TPj}, Q_{WPj}, Q_{Pij}, Q_{PJj}, $Q_{PBj} \geq 0$ (10)

Table 2. Description for elements of mathematical model

Element	Description
$\sum_{j \in J} F_l \times Z_l$	Fixed cost for opening a maintenance center
$\sum_{j \in J} (A_j + BH_j \times t_{HWj}) \times Q_{HWj}$	Total cost for goods from factory in HCMC to warehouse
$\sum_{j \in J} (A_j + CT_j \times t_{TPj}) \times Q_{TPj}$	Total cost for goods from Factory (TG) to Cat Lai Port
$\sum_{j \in J} (A_j + CW_j \times t_{WPj}) \times Q_{WPj}$	Total cost for goods from warehouse to Cat Lai Port
$\sum_{j \in J} (A_j + DI_j \times t_{PIj}) \times Q_{PIj}$	Total cost for goods from Cat Lai Port to IKEA
$\sum_{j \in J} (A_j + DJ_j \times t_{PJj}) \times Q_{PJj}$	Total cost for goods from Cat Lai Port to JYSK
$\sum_{j \in J} (A_j + DB_j \times t_{PBj}) \times Q_{PBj}$	Total cost for goods from Cat Lai Port to Bettenwelt
$\sum_{l \in L} \sum_{j \in J} EI_j \times t_{IC} \times Y_{ICj}$	Total cost for goods from market demand IKEA to maintenance center l
$\sum_{l \in L} \sum_{j \in J} EJ_j \times t_{JC} \times Y_{JCj}$	Total cost for goods from market demand JYSK to maintenance center l
$\sum_{l \in L} \sum_{j \in J} EB_j \times t_{BC} \times Y_{BCj}$	Total cost for goods from market demand Bettenwelt to maintenance center l
$\sum_{l \in L} \sum_{j \in J} (-GI_j + EI_j \times t_{ICj}) \times TI_{lrj}$	Total cost for goods from maintenance center l back to IKEA.
$\sum_{l \in L} \sum_{j \in J} (-GJ_j + EJ_j \times t_{JCj}) \times TJ_{lrj}$	Total cost for goods from maintenance center l back to IKEA.
$\sum_{l \in L} \sum_{j \in J} (-GB_j + EB_j \times t_{BCj}) \times TB_{lrj}$	Total cost for goods from maintenance center l back to IKEA.

The objective function is minimization of the total cost. The primary parts show the fixed costs of opening maintenance centers, respectively. The second part represents the production and transportation costs of product j from factory m to warehouse. The third part represents the production and transportation costs of product j from warehouse to market demand k. The fourth part represents the production and transportation costs of product j from market demands k to maintenance center l. The ultimate parts associated with cost of recovery product.

Constraint (1) indicates that quantity of product j produced by factory so as to transfer to warehouse is equal to quantity of product j from warehouse to market demand k. Constraint (2) ensures that the total number of product j for every market demand is equal or greater than the demand. Constraint (3) is about number of products from maintenance center to plant and number of products produced at plant to transfer to distribution center must be smaller than or equal the capacity of plant. Constraint (5) shows the capacity constraint of maintenance center. Constraint (6), (7), (8) shows the returned product. Constraint (9) ensures the binary nature of decision variables while constraint (10) preserves the non-negativity restriction on the decision variables.

4. Results:

4.1. Hierarchical clustering results:

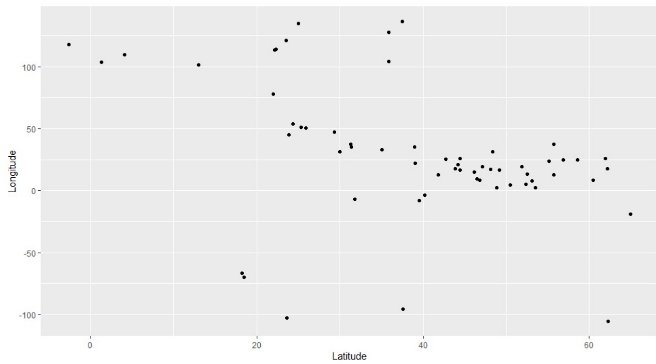


Figure 6: Store distribution

The stores of partners are distributed worldwide as below. By the means of using Hierachical test, the optimal number of clusters are shown as follows:

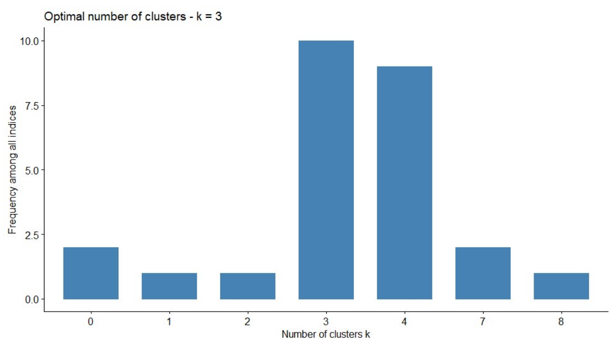


Figure 7: Optimal number of clusters, k=3

According to the results of frequency test, the most optimal number of clusters is k=3 which almost reaches 10.0 and the local optimal one is 4 clusters which is around 9.0.

To get a closer look at the results, several tests have been conducted to examine as follow:

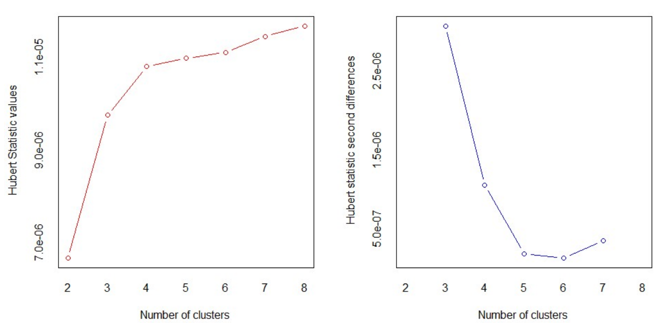


Figure 8: Hubert Statistic test determining the best number of clusters in the simulated data set

Figure 8 indicates that three clusters works better than four clusters as there is a dramatical fall in the results from k=3 to k=4 the Hubert statistics test for determining the best number of clusters in the simulated data set. While the other k does not make much difference from each other.

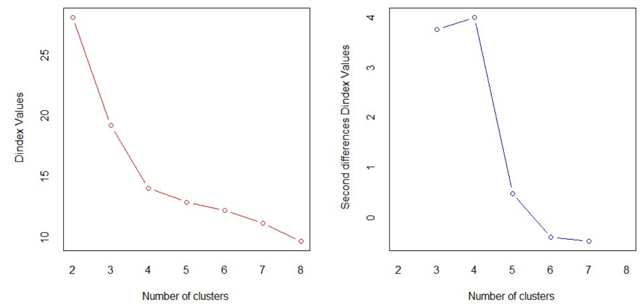


Figure 9: Dindex test for determining the best number of clusters in the simulated data set

On the other hand, figure 9 shows that the value reach the peak at k=4 rather than the local optimal k=3 when conducting the Dindex test for determining the best number of clusters in the simulated data set.

Both tests indicate the results of three clusters and four clusters without much difference. So, silhouette test is then conducted to reach the better conclusion. The 3-clustering group is distributed as the dendrogram follows:

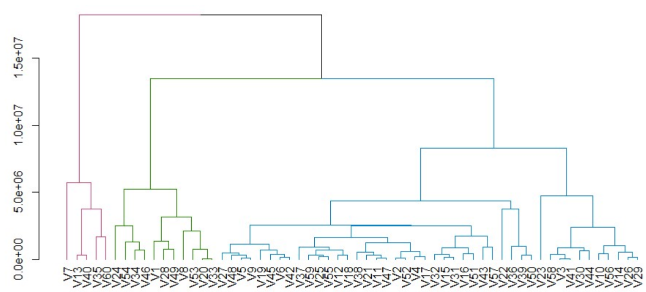


Figure 10. Three-clustering dendrogram

Which can also be distributed on the map as follows:

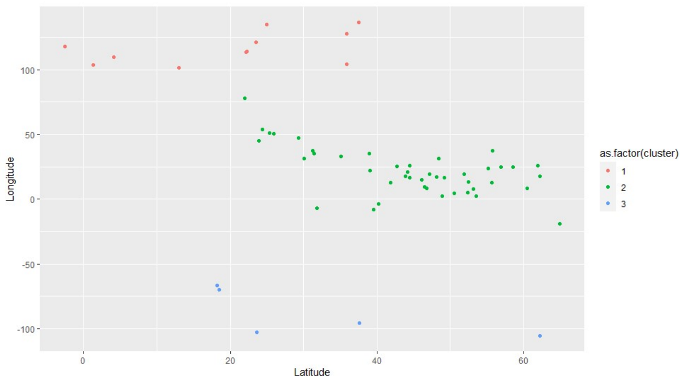


Figure 11. Three clusters distributed by colors on map.

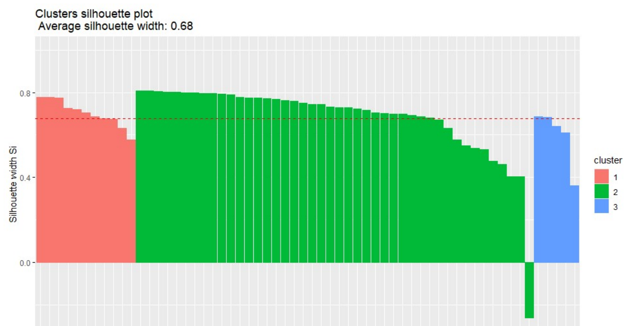


Figure 12: Cluster silhouette plot for three clusters

The cluster silhouette plot for three clusters shows a good sight with the average silhoutte width of 0.68 which is over 0.50. However, according to the thickness of the second cluster, it seems like the facility location might be overload. Moreover, there is a negative value appears in the second cluster which outlines that the second cluster is not well-cluster. Since then, in this case, four clusters are better for the case study. With four clusters, the cluster silhoutte plot are showned as follows:

Figure 13: Cluster silhoutte plot for four clusters

According to this Clusters silhouttes plot, four clusters group is better due to the limited capacity of the facility location for each cluster. In addition, the average silhoutte width is 0.58 which is still over 0.50 and there is no negative value in all clusters. The four clusters are distributed on the map as:

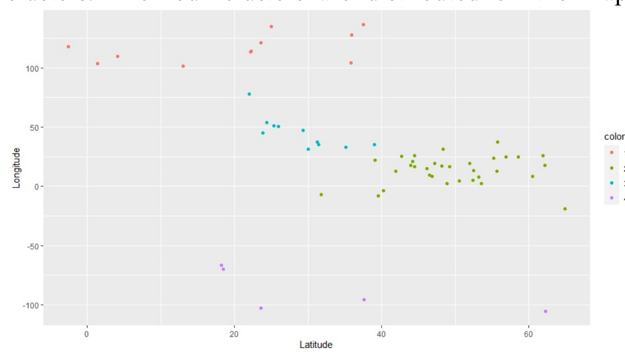


Figure 14: Four clusters distributed by colors on map.

The group is distributed as the dendrogram as follows:

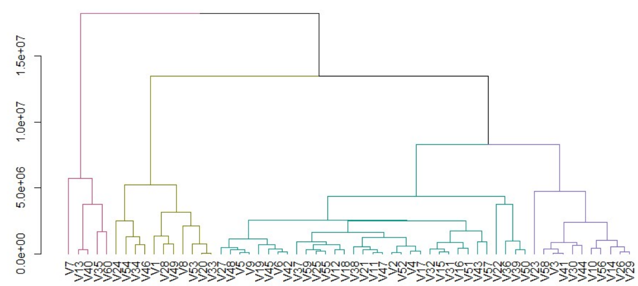


Figure 15: Four-clustering dendrogram

According to the distributed clusters as the dendrogram has shown, the clustering is as below with their calculated value of COG and the initial facility location:

Table 13: Four clusters with COG and initial facility location for maintenance centers

Cluster No.	Market	x-coordinate of maintenance center	y-coordinate of maintenance center	Maintenace center address
1	Canada, Dominican Republic, Mexico, Puerto Rico, United States	40.25	-95.03	Hughes Township, Nodaway County, Missouri, Missouri, USA
2	Australia, China, Hong Kong, Indonesia, Japan, Macau, Malaysia, Singapore, South Korea, Taiwan, Thailand	28.75	114.47	Xiu Shui, Jiujiang, Jiangxi, China
3	Austria, Belgium, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Morocco, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, The Netherlands, Ukraine, United Kingdom	51.23	14.22	Panschwitz-Kuckau, Germany
4	Bahrain, Cyprus, Egypt, India, Israel, Jordan, Kuwait, Qatar, Saudi Arabia, Turkey, United Arab Emirates	30.75	38.44	Al Qurayyat Saudi Arabia

The application of Hierachical slustering algorithm and COG has led to four clusters with four facility location which are:

- Cluster 1: Hughes Township, Nodaway County, Missouri, Missouri, USA
- Cluster 2: Xiu Shui, Jiujiang, Jiangxi, China
- Cluster 3: Panschwitz-Kuckau, Germany
- Cluster 4: Al Qurayyat Saudi Arabia

5. Conclusion and Recommendation

5.1. Summary of results:

Prior work has documented the effectiveness of reverse logistics in improving the environement states and reducing cost for the organisations; Kannan et al (9) for example with carbon footprint based reverse logistics network design model which helps to decrease the amount of carbon dioxide while minimizing costs for the organisations. However, these studies have either been using clustering in solving a facility location problem for reverse logistics systems. In this study a hierachical clustering algorithm is used to find the near optimal solution of the proposed model and the number of optimal clusters for the facility location. Meanwhile a MILP model is developed for multistage reverse logistics network design for the exact purpose of minimizing the total costs including fixed opening costs, transportation costs, saving costs.

The sum of everything which has been stated so far, four initial facility location for four cluster of is proposed to solve facility location problem for reverse logistics systems among the lines of: (Cluster 1) Hughes Township, Nodaway County, Missouri, Missouri, USA; (Cluster 2) Xiu Shui, Jiujiang, Jiangxi, China; (Cluster 3) Panschwitz-Kuckau, Germany; (Cluster 4) Al Qurayyat Saudi Arabia.

By applying reverse logistics networks, Scancom does not only benefit from improving customer satisfaction but also impression from customers. As the fact that reverse logistics supports customer services by getting defective products back into the reverse logistics line so it can be then reused, remanufactured, or properly disposed. Additionally, returning the used products that no longer in use from customers back to the reverse flow of logistics also helps to save costs for raw materials and reduce environmental pollution, especially deforestation. Finally, minimizing the total cost is also be accounted as the objectives of this research so that reverse logistics systems can be available.

5.2. Recommendation:

Firstly, in the hierarchical clustering models a distance matrix is conducted and the used method to find the distance is Euclidean distances which is common in clustering problems, but it still has some drawback and there is room that it can be replaced to another method for future works.

Secondly, besides minimizing the costs, the objective functions should include some variables such as “responsiveness”, “robustness” or limited amount of returned produced under the reverse logistics networks circumstances. Therefore, the applied MILP models should be upgraded to achieve the efficiency of the models and the networks.

Finally, there are five basic types of clustering, namely: hierarchical methods, partitioning techniques, density, or mode seeking techniques, clumping techniques and other methods not falling into the other categories. Some of the suitable for the location-based strategy can be counted as: DBSCAN, OPTICS, PAM and CLARA clustering can be considered to solve the facility location problems.

Reference

Jayaraman, V., & Ross, A. (2003). *A simulated annealing methodology to distribution network design and management*. European Journal of Operational Research, 144(3), 629-645.

Jimenez, G., Santos, G., Félix, M., Hernández, H., and Rondón, C. (2019). *Good Practices and Trends in Reverse Logistics in the plastic products manufacturing industry*. Procedia Manufacturing, 41, 367-374.

Rogers, D. S., & Tibben-Lembke, R. S. (1999). *Going backwards: reverse logistics trends and practices (Vol. 2)*. Pittsburgh, PA: Reverse Logistics Executive Council.

Melo, M. T., Nickel, S., & Saldanha-Da-Gama, F. (2009). *Facility location and supply chain management–A review.* European journal of operational research, 196(2), 401-412.

Lieckens, K., & Vandaele, N. (2007). *Reverse logistics network design with stochastic lead times*. Computers & Operations Research, 34(2), 395-416.

Rogers, D. S., Melamed, B., & Lembke, R. S. (2012). *Modeling and analysis of reverse logistics*. Journal of Business Logistics, 33(2), 107-117.

Pishvaee, M. S., Kianfar, K., & Karimi, B. (2010). *Reverse logistics network design using simulated annealing*. The International Journal of Advanced Manufacturing Technology, 47(1-4), 269-281.

Pishvaee, M., Farahani, R., & Dullaert, W. (2010). A memetic algorithm for bi-objective integrated forward/reverse logistics network design. *Computers & Operations Research*, *37*(6), 1100-1112. doi: 10.1016/j.cor.2009.09.018

Kannan, D., Diabat, A., Alrefaei, M., Govindan, K., & Yong, G. (2012). *A carbon footprint based reverse logistics network design model*. Resources, conservation and recycling, 67, 75-79.

Kuşakci, A. O., Ayvaz, B., Cin, E., & Aydın, N. (2019). *Optimization of reverse logistics network of End of Life Vehicles under fuzzy supply: A case study for Istanbul Metropolitan Area*. Journal of cleaner production, 215, 1036-1051.

Azizi, V., Hu, G., & Mokari, M. (2020). *A two-stage stochastic programming model for multi-period reverse logistics network design with lot-sizing*. Computers & Industrial Engineering, 143, 106397.

Kargar, S., Pourmehdi, M., & Paydar, M. M. (2020). *Reverse logistics network design for medical waste management in the epidemic outbreak of the novel coronavirus (COVID-19)*. Science of the Total Environment, 746, 141183.

Liao, T. Y. (2018). *Reverse logistics network design for product recovery and remanufacturing*. Applied Mathematical Modelling, 60, 145-163.

Parchami Afra, A., & Behnamian, J. (2021). Lagrangian heuristic algorithm for green multi-product production routing problem with reverse logistics and remanufacturing. *Journal Of Manufacturing Systems*, *58*, 33-43. doi: 10.1016/j.jmsy.2020.11.013

Trochu, J., Chaabane, A., & Ouhimmou, M. (2018). *Reverse logistics network redesign under uncertainty for wood waste in the CRD industry*. Resources, Conservation and Recycling, 128, 32-47.

Kanungo, T., Mount, D., Piatko, C., Silverman, R., Wu, A., "*An efficient k-means clustering algorithm: analysis and implementation*" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.7, pp. 887-892, (2002).

Krishna, K. and Murty, M. Narasimha, "*Genetic k-means algorithm*" IEEE Transactions on Systems Man And Cybernetics-Part B: Cybernetics, Vol.29, No.3, pp. 433-439, (1 999)

Camastra, F. and Verri, A., "*A novel kernel method for clustering*" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.5, pp. 801-804, (2005).

Cheung, Y.M., "*K-means: a new generalized k-means clustering algorithm*" Pattern Recognition Letters, Vol.24, No.15, pp. 2883-2893, (2003).

Tsai, C.A., Lee, T.C., Ho, I.C., Yang, U.C., Chen, C.H., Chen, J.J., "*Multi-class clustering and prediction in the analysis of microarray data*" Mathematical Biosciences, Vol.193, Issue 1, pp. 79-100, (2005).

Qin, J., Lewis, D.P., Noble, W.S., "*Kernel hierarchical gene clustering from microarray expression data*" Bioinformatics, Vol. 19, No.16, pp. 2097-2104, (2003).

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). “*Quantum machine learning”*. Nature, 549(7671), 195-202.

Huang, B. (2017). "*Comprehensive geographic information systems"*. Elsevier.

Sheu, J. B., & Lin, A. Y. S. (2012). *Hierarchical facility network planning model for global logistics network configurations*. Applied Mathematical Modelling, 36(7), 3053-3066.

DESIGN SUPPLY CHAIN NETWORK FOR PANGASIUS FEED TO THE MEKONG DELTA - WOOSUNG VINA COMPANY CASE STUDY

Nguyen Le Cat Quynh[1], Duong Vo Nhi Anh[2]

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: [1]nlcquynh.0405@gmail.com, [2]dvnanh@hcmiu.edu.vn

Abstract: High transportation cost is now a negative effect on the total cost of the supply chain, motivating researchers to find a better solution to reduce the cost. Overstocking also leads to the waste of products. This paper suggests the optimization of the supply chain network with backhaul. The proposed mathematical model considered the selection of warehouses, backhauls from distribution centers to suppliers before vehicles going back to factories, and the number of routes assigned for each vehicle size. The proposed model is then verified using IBM ILOG CPLEX Optimizer with a set of customers in the Mekong Delta by roadway to help the organization control and further implement in the real working situation.

Keywords: Supply Chain Network, Optimization, Backhaul, Transportation, Warehouse Location Problem

1. Introduction

1.1. Background

Supply chain management (SCM) chain can be defined as all the related activities on planning system from upstream to downstream consisting of suppliers, producers, manufacturers, warehouses, distribution centers, retailers, and customers. These activities establish distribution channels, amount of materials and items to consume, produce, and ship from suppliers to customers, location, capacity, and type of plants, warehouses, and distribution centers are determined to minimize the total cost of the product distribution system.

A cost of the supply chain comprises fixed cost and variable cost. Fixed costs are the cost for operating facilities that are not related to the quantities of parts or products passing through the distribution network. For variable costs, there are transportation costs, storage costs, and operation costs that are directly varied with the quantities of products or parts passing through the network.

Transportation cost is one of the critical costs of the distribution system for any company. Backhaul is a part of the transportation cost that causes the company to waste money when vehicles return to their original location with empty vehicles.

The current supply chain of the Woosung Vina Company has three problems:

- The output of the factory does not meet 15-20% of the demand for pangasius feed in peak season, from January to April.
- The logistics cost, especially transportation cost is higher than competitors at about 12% in 2020.
- The monthly warehouse rental cost of the current warehouse in My Tho City, Tien Giang is high, about 80 billion VND for a 1200m2 area. Moreover, the warehouse is located in a residential area and is under legal pressure because of the nasty smell and smog. In addition, the warehouse leasing contract is about to expire after 5 years (from 2017 to 2022) and the company is looking for investing in another smaller warehouse location as the distance between the warehouses and the location of end-buyers are supposed to be more reasonable than the current status.

To tackle the above problems, this research will provide:

- A material requirement planning needs to be developed for better material purchasing planning and stabilizing the market.
- Minimize stock out in peak season by assigning a safety stock.
- Providing a suggestion for new smaller warehouses location which are nearer to customers to improve customer service.

- Improve the transportation performance for the company up to 39% by using backhaul in transportation.
- Create a national supply chain network of feed products in Woosung Vina in order to minimize the total logistics cost includes transportation cost based on selecting the right vehicles.

The expected output is that the company can easily manage the delivery process and warehouse inventory information. This particular thesis is going to propose a new division networking for Woosung Vina pangasius feed to help the growers in Woosung manage the growing transporting cost. Moreover, a redesign of supply chain network needs to be done in order to improve the quality of customer service, create a stable profit model for customers, and better optimization of cost savings.

2. Literature Review

2.1. Transportation in Supply Chain

Along with the increased need of improving the sustainability of processes and optimizing their efficiency, the demand of delivering the right product, in the right quantity to the right place at the right time also puts pressure on businesses. Transportation is one of the most fundamental and structuring activities within a supply chain. Transportation plays a vital role in providing the critical linkages between the supply chain networks within businesses located both domestically and internationally. It is responsible for a big share of logistics costs, around one to two-third (Ballou, 2006). For this importance, the pursuit of optimization opportunities in this area is a priority for companies that can contribute to improved supply chain efficiency.

With the importance of transportation in a supply chain, it cannot be managed in an isolated manner, as the options of the mode and the type of transportation have a significant impact on the cost structure and in the company's ability to react to the market. Only good coordination between each component would bring the benefits to a maximum (Tseng, 2005).

2.2. Distribution network problems

Refer to the dedication of places as well as the phrase allocation refer to assigning tasks, or perhaps to creation sites, distribution centers as well as customers. Osman ALP (2003) proposed in the paper a brand-new genetic algorithm for the issue of site and places which come in the picking of the very best locations of P sites to deliver locations of N site in the foundation of minimization the distance in-between areas.

Rami Musa et al (2010) proposed in the paper of his novel algorithm to resolve the issue of the commuter routes of the cross-docking distribution network on the foundation of reducing the expense of living from established I of Suppliers to a pair of J consumers within the transit of a set C of cross-docking distribution centers. The Ant Colony Algorithm was utilized to resolve the issue.

Partners and fethi Boudari (2013) published in their newspaper Reference a unit to discover the assignment of the client of a merchant cluster in a community in what the buyers can get back their merchandise to the merchant. This paper focuses primarily on the poultry market of the community of Tlemcem, Algeria.

2.3. Design supply chain network

The network design issue consists of choosing the facilities to open to be able to conduct a division with minimized costs but still have a high customer service level. The design issues of a supply chain are usually hard due to the connection with economic, social, and organizational sectors. Therefore, the design of an effective and productive logistic community has attracted the soaring interest of both the scientific and industrial communities.

Fan Wang (2011) suggested the modeling and solving of a supply chain design for annual cost minimization while considering environmental effects. The supply chain cost elements considered in this paper are transportation cost, holding cost, open facility fixed cost, and variable cost. The second objective is to minimize the amount of NO_2, CO, and volatile organic gases. They utilize a memetic algorithm in combination with the Taguchi method to solve this model, and they also proposed a novel decoding method and priority-based algorithm for coding the solution chromosome that has been examined against the hybrid genetic Taguchi algorithm (GATA). And the proposed method can effectively provide better results. Jamshidi (2012) also demonstrated a green operations supply chain management considers three environmental levels for facilities and only one environmental level can be set for any opening facility. they solve the multiple-objective model with each objective function separately to get the objective value $\mu 1$ and $\mu 2$ and the solution vector $x1^*$ and $x2^*$ corresponding to OBJ2 and OBJ1, respectively. After that, the author used the normalized design objective space, and join the two objective values $\mu 1$ and $\mu 2$ with a line called Utopia line to get the Pareto optimal set for a multi-objective final solution. While Dzupire, N. C., & Nkansah-Gyekye, Y. (2014) using genetic algorithm to solve the problem, Altiparmak (2006) proposed multi-objective with weight and used genetic algorithms to find the set of Pareto-optimal solutions for SCN design problem. Then, the proposed solution procedure was compared with simulated annealing according to quality of Pareto-optimal solutions in the second stage.

Wei-Chang Yeh (2005) developed an efficient hybrid heuristic algorithm (HHA) by combining a greedy method (GM), the linear programming technique (LP) and three local search methods (LSMs). All the mentioned methods were implemented into the binary variable vector and must be performed in series until the objective function value is unchanged.

Kocaoglu's (2020) research introduces three distribution strategies: direct shipment, cross-docking, and milk-run, and the authors developed a mathematical model to determines the proper distribution strategies and optimal routes. The objective of the model is to minimize the total of (direct shipment cost + milk run cost + cross-docking cost). First, they use GA and randomly assign the distribution strategies to customers. After that, they will apply a modified savings algorithm if the method is equal to milk run, direct procedure if the method is equal to direct shipment, and cross-docking procedure if the method is equal to cross-docking. After that, they calculate the total distribution costs and rank the costs in descending order. The next step is creating a new population by applying GA through selection, crossover, mutation, and placement stages. If the population size is equal to 100, finish forming a new offspring. Then, Replace-Apply the elitist strategy and use the newly generated population. Terminate algorithm if iteration number is equal maximum iteration number, finish algorithm, and return the best solution. Meanwhile, Boonmee (2015), vehicles left from the factory to distribute products to distribution center and then vehicles went to pick up raw material at a supplier before went back to the factory. Backhaul was reduced when introducing the routes from suppliers to plants. The mathematical model considered the selection of the number of appropriate distribution centers, the size of vehicles, and the number of traveling times by each vehicle.

3. Methodology

3.1. Conceptual Design Description

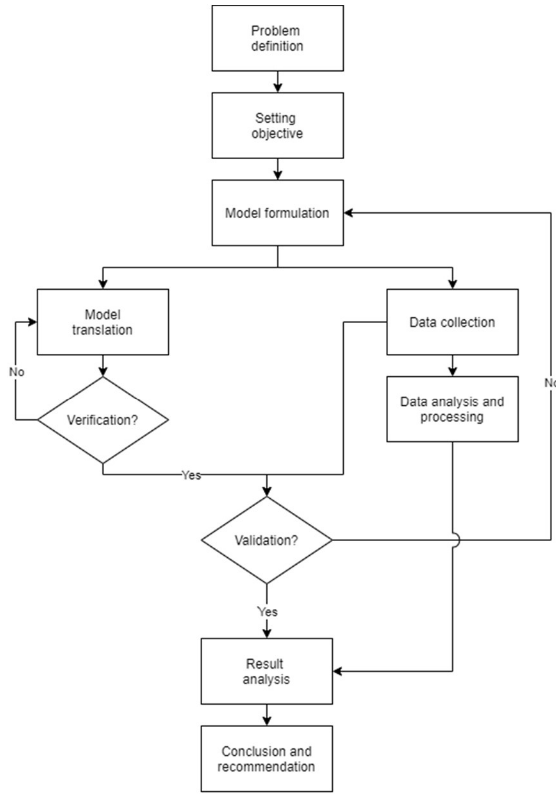


Figure 1: Conceptual design for this study

3.1.1. Problem definition

This information is indicated in the introduction chapter. In this section, the current problems of the Woosung Vina supply chain are defined. Finding the paper is very important to the thesis, based on the related paper, the methods can be identified and implied.

3.1.2. Setting objective

In this stage, the value chain of pangasius at Woosung Vina will be analyzed. Based on that, the study will propose a more suitable supply chain network for pangasius farmers.

3.1.3. Model formulation

In this stage, the mathematical model is built based on the objective is minimizing the total transportation cost through selecting a suitable vehicle for different demands of pangasius feed products. To check the validity of the proposed theoretical system, a simulation model with the conceptual model, assumed data input is built and run to have the final result.

3.1.4. Model translation

ILOG IBM CPLEX software will be applied to solve Mixed Integer Programming for designing supply chain network by inputting the mathematical into coding.

3.1.5. Verify the model

This process ensures that if the model behaves as intended and is accurate.

3.1.6. Data collecting

Based on the list of parameters, essential data are collected.

3.1.7. Data analysis and processing

The next step is data modification, then necessary data is verified and input into the model to solve it. There will be two data sets, one is the historical data set for the proposed supply chain network to validate the model. The collected data, along with building model, code is put into CPLEX to find the optimal solution for proposed supply chain optimization.

3.1.8. Model validation

After building the suitable model, data is applied to be made to check whether not these results come as an improvement comparing to the current situation of the company. If they match the expectation, move to the next step. If not, a modification will be made to obtain the final optimization model and results.

3.1.9. Result analysis

Input the forecast data set to the final model to come up with the result.

3.1.10. Conclusion and recommendation

According to the results from the previous step, there will be a suggestion for the business on how to implement to improve the business. There should also be a discussion on how the proposed model can be upgraded.

3.2. Modelling

3.2.1. Assumptions

- All used parameters are known, constant and deterministic.
- Inventory of raw materials were not allowed.
- Time unit is monthly.
- Vehicles left from factory or warehouse to distribute finished products and then go to pick up raw materials at the suppliers before going back to the origin.
- Linehaul vehicles from warehouse to customers will then collab with external parties in order to transport goods from customers' province to warehouse's province.
- Two considered warehouses (An Giang, Vinh Long) was conducted by using Clustering model in data processing phase.
- A safety stock is assigned to cover stock-out.

3.2.2. Index

m index of materials, m = 1 ... M
f index of feed products, f = 1 ... F
s index of suppliers, s = 1 ... S
p index of plants, p = 1 ... P
w index of warehouses, w = 1 ... W
c index of customers, c = 1 ... C
v index of vehicles, v = 1 ... V
t index of periods, t = 0 ... T

3.2.3. Parameters

$\mathbf{DM_{xy}}$ distance from location x to location y
$\mathbf{a_{spv}}$ transportation cost from supplier s to plants p by vehicle v
$\mathbf{b_{pwv}}$ transportation cost from plant p to warehouse w by vehicle v
$\mathbf{c_{wcv}}$ transportation cost from warehouse w to customer c by vehicle v
$\mathbf{d_{pcv}}$ transportation cost from plant p to customer c by vehicle v
$\mathbf{price_m}$ purchase price of material m
$\mathbf{ope_f}$ operations cost to produc1 unit feed f
$\mathbf{hold_f}$ unit inventory cost of product f
$\mathbf{fix_w}$ Annual fixed cost for investing warehouse w
$\mathbf{cap\_s_s}$ capacity of supplier s
$\mathbf{cap\_p_p}$ production capacity of plant
$\mathbf{cap\_wh_w}$ capacity of the warehouse w
$\mathbf{cap\_v_v}$ capacity of vehicle type v
$\mathbf{D_{fct}}$ demand for product f in customer c in period t
$\mathbf{M_{fm}}$ amount of material m needed to produce 1 unit of feed f
$\mathbf{initial_f}$ Initial inventory level of product f
$\mathbf{B_{ct}}$ Backhaul demand of customer c in period t

Transportation cost parameters are considered by distance matrix (DM_{xy}) and unit cost in VND/trip by truck as follow

Table 1: Transportation parameter list

Parameter	1.25-ton truck	8-ton truck
a_{spv}	1.25cost*DM_{sp}	8cost*DM_{sp}
b_{pwv}	1.25cost*DM_{wc}	8cost*DM_{pw}
c_{wcv}	1.25cost*DM_{wc}	8cost*DM_{wc}

3.2.4. Decision variables

$\mathbf{inv_{ft}}$ inventory level of product f in period t
$\mathbf{Xa_{mspvt}}$ the amount of material *m* transported from supplier *s* to plant *p* by vehicle *v* in period *t*

Xb_{fpwvt} the amount of product f transported from plant p to warehouse w by vehicle v in period t

Xc_{fwcvt} the amount of product f transported from warehouse w to customer c by vehicle v in period t

Va_{spvt} the number of trips by vehicle v travelled from supplier s to plant p in period t

Vb_{pwvt} the number of trips of vehicle v travelled from plant p to warehouse w in period t

Vc_{wcvt} the number of trips vehicle v travelled from warehouse w to customer c in period t

$V1_{wsvt}$ the number of vehicle v travelled from warehouse w to supplier s in period t

$V2_{cvt}$ number of backhaul routes from customer c by vehicle v in period t

3.2.5. Objective function

$$\begin{aligned}&\sum_{s}\sum_{p}\sum_{v}\sum_{t} a_{spv}\,Va_{spvt} + \sum_{p}\sum_{w}\sum_{v}\sum_{t} b_{pwv}\,Vb_{pwvt} + 2\sum_{w}\sum_{c}\sum_{v}\sum_{t} c_{wcv}\,Vc_{wcvt} + \sum_{w}\sum_{s}\sum_{v}\sum_{t} d_{wsv}\,V1_{wsvt} \\ &+ \sum_{m}\sum_{s}\sum_{p}\sum_{v}\sum_{t} price_m\,Xa_{mspvt} + \sum_{w} fix_w + \sum_{f}\sum_{t} hold_f\,inv_{ft} + \sum_{f}\sum_{p}\sum_{w}\sum_{v}\sum_{t} Xb_{fpwvt}\,ope_f - \sum_{c}\sum_{t} revenue_c * B_{ct}\end{aligned} \quad (1)$$

3.2.6. Constraints

$$\sum_{s}\sum_{p}\sum_{v} Xa_{mspvt} = \sum_{f} M_{fm}\sum_{f}\sum_{p}\sum_{w}\sum_{v} Xb_{fpwvt}, \forall m,t \quad (2)$$

$$\sum_{m}\sum_{v}\sum_{p} Xa_{mspvt} \le cap\_s_s, \forall s,t \quad (3)$$

$$inv_{f0} = initial_f \;\forall f \quad (4)$$

$$\sum_{m}\sum_{s}\sum_{v} Xa_{mspvt} - \sum_{f}\sum_{w}\sum_{v} Xb_{fpwvt} \ge 0, \forall p,t \quad (5)$$

$$\sum_{p}\sum_{v} Xb_{fpwvt} - \sum_{c}\sum_{v} Xc_{fwcvt} \ge 0 \;\forall f,w,t \quad (6)$$

$$\sum_{f}\sum_{w}\sum_{v} Xb_{fpwvt} \le cap\_p_p, \forall p,t \quad (7)$$

$$\sum_{f}\sum_{p}\sum_{v} Xb_{fpwvt} \le cap\_wh_w, \forall w,t \quad (8)$$

$$\sum_{w}\sum_{v} Xc_{fwcvt} = D_{fct}, \forall f,c,t \quad (9)$$

$$inv\_pro_{ft} = inv\_pro_{f,t-1} + \sum_{p}\sum_{v} Xb_{fpwvt} - \sum_{c}\sum_{v} Xc_{fwcvt}, \forall f,w,t \quad (10)$$

$$\sum_{c}\sum_{v} Xc_{fwcvt} \le \sum_{p}\sum_{v} Xb_{fpwvt} + inv_{f,t-1} \;\forall f,w,t=1 \quad (11)$$

$$inv\_pro_{ft} \ge safety_f, \forall f,t \quad (12)$$

$$\sum_{v} Va_{spvt}\,cap\_v_v \ge \sum_{m}\sum_{v} Xa_{mspvt}, \forall s,p,t \quad (13)$$

$$\sum_{v} Vb_{pwvt}\,cap\_v_v \ge \sum_{f}\sum_{v} Xb_{fpwvt}, \forall p,w,t \quad (14)$$

$$\sum_{v} Vc_{wcvt}\,cap\_v_v \ge \sum_{f}\sum_{v} Xc_{fwcvt}, \forall w,c,t \quad (15)$$

$$\sum_{p} Va_{pwvt} = \sum_{w} V1_{wsvt}, \forall s,v,t \quad (16)$$

$$\sum_{s} Va_{pwvt} = \sum_{w} Vb_{pwvt}, \forall p,v,t \quad (17)$$

$$\sum_{v} V2_{cvt} * cap\_v_v \ge backhaul_{ct}, \forall c,t \quad (18)$$

$$V2_{cvt} \le \sum_{w} Vc_{wcvt} \;\forall c,v,t \quad (19)$$

Objective function (1) is to minimize to minimize all existing costs including transportation costs, fixed investing cost, variable operational and purchasing cost with various indexes minus the revenue if backhaul demand is fulfilled. While constraint (2) calculates the amount of material will be transported from supplier to plant satisfy the quantity of product that will ship to the warehouses, (3) ensure that the total number of raw materials m shipped from supplier s must satisfy suppliers' capacity. Constraint (4) represent the assignment of initial inventory to the inventory level of product at period 0. (5) and (6) gives the inbound and outbound of plants and warehouses respectively. (7) is the plant production capacity constraint while (8) is the limit of warehouse capacity constraint. Constraint (9) gives the satisfaction of customer demands for the products. Constraint (10) is to calculate the inventory level of each product at the end of each period and (11) impose the output of the warehouses in period 1. Constraint (12) restrict the inventory level at the end of each period. (13)-(15) ensure the total amount of transported products and material of each time should not exceed the capacity of each vehicle capacity. Equations (16) and (17) are the constraint to ensure that the number of trips going into each node should be equal to the number of trips departure from that node. Constraint (18) calculates the number of vehicles serve backhaul demand and (19) makes sure backhaul trips must not exceed number of trips going into each customer node.

3.3. Data Collection

The data was gathered from the database of Woosung Vina Company followed their permission and privacy policies, including:

- Location of suppliers, plant, customers
- Customer demand
- Ingredient formulation
- Unit travelling cost
- Etc.

The data is provided by the estimated sale of pangasius feed in 2021 of ten Mekong Delta markets: Dong Thap, An Giang, Ben Tre, Can Tho, Vinh Long, Hau Giang, Tien Giang, Soc Trang, Tra Vinh, Long An. The crude protein content (CP) at 28% is consumed by 5 – 30 gram juvenile fish, while CP determined 26% is mainly used, about 70% to 80% in the feeding program, from 30 – 35 gram adult fish until 1 – 1.2 kilogram selling fish.

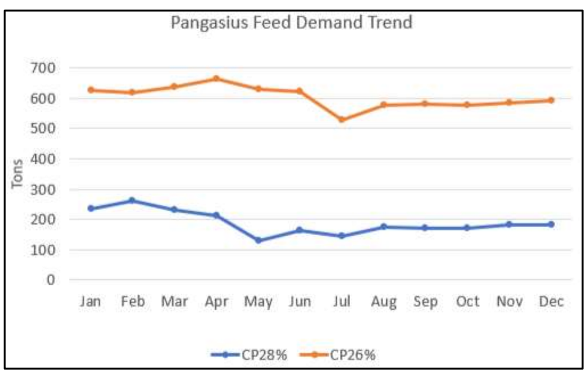


Figure 2: Pangasius feeds demand trend

4. Result Analysis

4.1. Data processing

4.1.1. Generating distance

One critical input for every supply chain network design problem is the distance matrix, which is basically the road distance between each node Generally, the distance processing phase will implement in three applications:

- ezGeocode, a process of converting addresses into geographic coordinates (like latitude and longitude), using Google Sheets add-in.
- Bing Maps REST Services Application Programming Interface (API) - a service that provides travel time and distances for a set of origins and destinations.
- Microsoft Excel and its add-in features Visual Basic for Applications (VBA).

The process of generating distance input:

- **Step 1**: Adding addresses into Google Sheets and using ezGeocode, to convert addresses into latitude and longitude.
- **Step 2:** Request for a personal API Key on Bing Maps.

API Key
AnCVRJTaJu8or0hmnaJCXvFw7Yl37yA3JIznBh D6V7jbvQ1TqxHl-NMlYE2Tymiz

- **Step 3:** Distance calculation

The function is called GetDistance coded in VBA, which will return the distance between two given coordinates in kilometers.

```
Public Function GetDistance(start As String, dest
As String, key As String)
Dim firstVal As String, secondVal As String,
lastVal As String
firstVal =
"https://dev.virtualearth.net/REST/v1/Routes/Dista
nceMatrix?origins="
secondVal = "&destinations="
lastVal = "&travelMode=driving&o=xml&key="
& key & "&distanceUnit=mi"
Set objHTTP =
CreateObject("MSXML2.ServerXMLHTTP")
Url = firstVal & start & secondVal & dest &
lastVal
objHTTP.Open "GET", Url, False
objHTTP.SetRequestHeader "User-Agent",
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT
5.0)"
objHTTP.Send ("")
GetDistance =
Round(Round(WorksheetFunction.FilterXML(obj
HTTP.ResponseText, "//TravelDistance"), 3) *
1.609, 0)
End Function
```

Origin	1	Input
Arrival	2	
Origin Coordinates	10.958135, 107.0276259	Function Output
Arrival Coordinates	10.2691019, 105.5319265	
Key	AnCVRJTaJu8or0hmnaJCXvFw7Y137yA3JIz nBhD6V7jbvQ1TqxHl-NMIYE2Tymiz	
Distance	235	

Figure 3: Result of Getting Distance VBA Function

After that, VBA is used continually to build up the DistanceMatrix function using the GetDistance function to all the cells in Distance Matrix.

```
Sub dismatrix()
For Each cell In Range("G2:U16")
Range("B1").Value = Cells(cell.Row, 6).Value
Range("B2").Value = Cells(1, cell.Column).Value
cell.Value = Range("B8").Value
Next cell
End Sub
```

4.1.2. Warehouses Location Problem

In this stage, Minitab Statistical Software is applied for generating Cluster Observations.

- **Step 1:** Adding latitude, longitude and total forecast demand of ten different regions into Minitab.
- **Step 2:** Testing Cluster Observations with the number of clusters is 1.

Table 3: One cluster analysis Minitab output

Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	9	93.3013	0.31381	5	6	5	2
2	8	77.5707	1.05072	8	9	8	2
3	7	72.7270	1.27763	4	5	4	3
4	6	71.2486	1.34689	3	7	3	2
5	5	69.2458	1.44071	1	2	1	2
6	4	65.2777	1.62660	3	4	3	5
7	3	48.4413	2.41532	3	8	3	7
8	2	39.3738	2.84010	3	10	3	8
9	1	0.0000	4.68460	1	3	1	10

Looking at the table shows the cluster:

- The similarity level decreases by increments of approximately 5 or less until step 7. The similarity decreases by around 39 (from 39.3738 to 0) at steps 8 and 9, when the number of clusters changes from 2 to 1.
- The distance between the joined clusters increases, first by approximately 0.8 or less. The distance increases by about 2 (from 2.84010 to 4.68460) at steps 8 and 9, when the number of clusters changes from 2 to 1.

It indicates that 2 clusters are reasonably sufficient for the final partition.

- **Step 3:** Apply 2 clusters for Cluster Observations

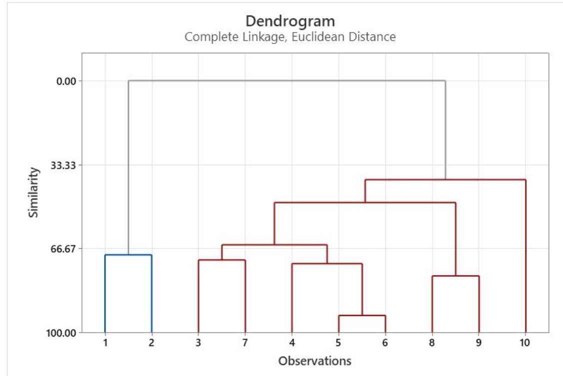


Figure 4: Three clusters dendrogram

According to the clusters dendrogram, the centroid of each cluster is calculated by Center of Gravity:

$$Latitude_{centroid} = \frac{\sum Lat_i Demand_i}{\sum Demand_i},$$

$$Longitude_{centroid} = \frac{\sum Long_i Demand_i}{\sum Demand_i}$$

Table 3: Two clusters with centroids coordinates

No	Name	Coor-dinates	De-mand	Clus-ter	Cent-roid Coor-dinates
1	Dong Thap	10.733, 105.367	2,994.37	1	10.5465, 105.4232
2	An Giang	10.328, 105.488	2,563.7		
3	Ben Tre	10.145, 106.229	1,286.01	2	10.1140, 105.9520
4	Can Tho	10.165, 105.640	1,284.56		
5	Vinh Long	9.973, 105.906	621.75		
6	Hau Giang	9.955, 105.864	310.99		
8	Soc Trang	9.724, 106.080	85.01		
9	Tra Vinh	9.913, 106.411	60.96		
7	Tien Giang	10.383, 106.058	213.07		
10	Long An	10.440, 106.599	58.6		

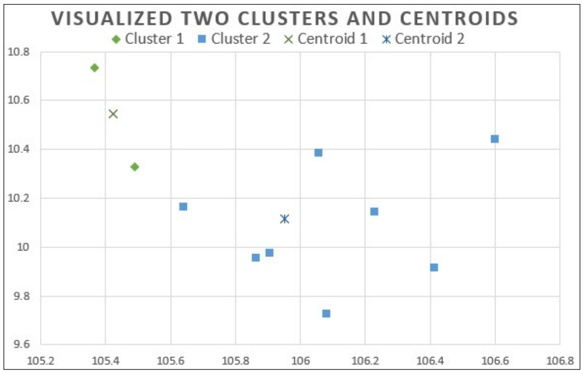


Figure 4: Visualize two clusters and centroids coordinates

The application of K-means clustering in Minitab and Center of Gravity has led to two centroid locations are:

Table 4: Centroid locations

Cent-roid	Coor-dinates	Location
1	10.54654, 105.42322	Long Dien A Commune, Cho Moi District, An Giang Province
2	10.11402, 105.9520	Phu Loc Commune, Tam Binh District, Vinh Long Province

However, all two above locations are the residential community that are not suitable to construct feed warehouse with a nasty smell and dust. As a result, there will be some adjust of an industrial zone nearby the result of Center of Gravity:

Table 5: Centroid adjusted locations

Cent-roid	Coor-dinates	Distance from COG	Adjusted Location
1	10.5708, 105.3583	9 km	**Tan Trung Industrial Zone**, Tan Trung Commune, Phu Tan District, An Giang
2	10.1713, 105.9365	12 km	**Hoa Phu Industrial Zone**, Hoa Phu Commune, Tam Binh District, Vinh Long

4.2. Mathematical Model Result

Based on the real data collected about the pangasius feed usage by farming system from Woosung Vina Company, the thesis solved the Optimization of Distribution Network with Backhaul by IBM ILOG CPLEX Optimization Studio 12.10. This leads to the optimal solution for pangasius feed supply chain network design, minimizing all cost related (transportation cost, feed producing cost, inventory cost, facility investing cost) minus the revenue of fulfilling customers backhaul demand has the figure of 144,672,238,672 VND. This section will depict how the supply chain network with backhaul achieve the optimal cost.

4.2.1. Costs from model output

Table 6: Summary of cost components in Backhaul

Cost	Value (VND)	Percentage
Material purchase	94,556,395,548 VND	65.36%
Operations	30,157,845,000 VND	20.85%
Transpor-tation	19,238,099,400 VND	13.30%
Warehouse rental	600,000,000 VND	0.41%
Holding product	119,898,725 VND	0.08%

4.2.2. Product quantity manufactured

To meet the demand requirement of ten different provinces with one-year period, the quantity of products that the company produces:

Figure 6: Product Quantity Manufactured

4.2.3. Inventory level

At the end of each period, the remaining quantity must greater than or equal the safety stock of each product (4 tons for CP28% product and 12 tons for CP26% product). The result of the proposed model introduce a fluctation in the inventory, The line chart below will illustrate the inventory level of each product:

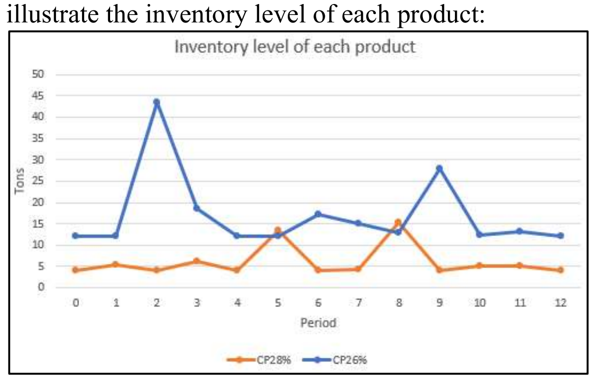


Figure 7: Inventory Level of each product

4.2.4. Transportation costs

By using two types of truck for flexibility in shipping activities, the result of each transportation cost is indicated below:

Table 7: Comparison transportation cost in two methods

Transportation	With backhaul	No backhaul
Suppliers to plant	6,485,370,000 VND	12,970,740,000 VND
Plant and warehouses	8,157,708,000 VND	16,315,416,000 VND
Warehouses to customers	3,009,960,000 VND	3,009,960,000 VND
Warehouses to suppliers	1,853,316,000 VND	-
Backhaul	-268,254,600 VND	-
Total cost	**19,238,099,400 VND**	**32,296,116,000 VND**
Savings 13,058,016,600 VND/year 40.4%		

4.3. Solution Development

As far as further improvement is concerned, Woosung Company should develop the long-term strategy by investing a plant specified in manufacturing aquaculture feed product in the Mekong Delta, where the majority of the farmers raising shrimp, pangasius, red tilapia, snakehead, carp, etc. and suppliers gather. This leads to the significant drop in the transportation cost from suppliers to plant and from plant to warehouse or even the company can distribute directly from plant to customers.

To test the result of this suggestion, the Center of Gravity for ten markets is implemented to find the location for the new plant:

Latitude	Longitude	Address
10.367637	105.64197	Vinh Thanh Commune, Lap Vo District, Dong Thap

However, the result of Center of Gravity leads to the location of residential area, which is not appropriate for the company to establish a factory. Instead, I recommend the company construct factory in Sa Dec Industrial Zone, where there are also various feedmills, such as: Cargill, Spotlight, Emivest and so on. In addition, The Sa Dec Industrial Zone is not far from the location given by the Center of Gravity (about 21km). As a result, the best location for the new Woosung Feedmill is:

Latitude	Longitude	Address
10.326804	105.75219	Zone C, Sa Dec Industrial Zone, Sa Dec City, Dong Thap

Thanks to the closer distance between plant, suppliers and warehouses, the transportation costs significant drop:

Table 8: Transportation costs comparison between Dong Nai and Dong Thap plant

	Dong Nai	Dong Thap
Suppliers to plant	6,485,370,000 VND	2,389,700,000VND
Plant and warehouses	8,157,708,000 VND	1,845,300,000 VND
Warehouses to customers	3,009,960,000 VND	3,009,960,000 VND
Warehouses to suppliers	1,853,316,000 VND	1,850,466,000 VND
Backhaul	-268,254,600 VND	-268,254,600 VND
Total cost	**19,238,099,400 VND**	**8,830,021,400 VND**

By investing a feedmill in Sa Dec Industrial Zone, there is a decrease of 10,408,078,000 VND in terms of annual transportation cost.

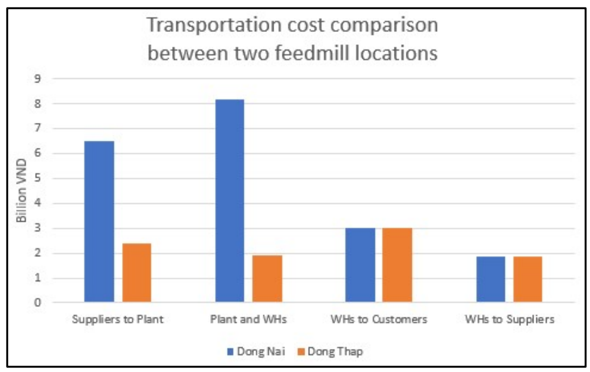


Figure 8: Transportation cost comparison between two factories location

According to the chart, investing a feedmill specifying in aquaculture feed product in Dong Thap assist the company take use of transportation costs. The shipping cost from suppliers to plant reduce by about 2.5 times, and that from plant to warehouses decrease nearly 5 times.

5. Conclusion and Recommendation

5.1. Conclusion

The effectiveness of backhaul has in enormous reducing the transportation cost up to 13 billion VND. This will assist the lower selling price of pangasius feed to the end users, thus increase the competitive advantages of Woosung Vietnam pangasius feed or increase profits for the company.

This study also applied clustering to solve the problem of optimal warehouse locations for customers in different locations and needs. In addition, a mathematical model of optimization supply chain network with backhaul using IBM ILOG CPLEX Optimization Studio provide the solution for the main goal of minimizing total logistics costs including fix operation costs, warehouses rental costs, holding costs, transportation costs and material purchasing costs. Three warehouses location from the result of three clusters is applied to solve location problem. This not only helps the company to control the quality of customer service in the current situation of rising transportation and material costs.

5.2. Limitation

In the first place, the warehouse locations problem is solved by using clustering model in which the Euclidean distances which is commonly used. Although it provided an appropriate solution in general, it may still have some flaws if it has to deal with more complicated problems. If there is more room, another method should be applied for future work.

Similarly, using CPLEX in finding the optimal solution of mathematical model is also have some limits. The mathematical model can be used to solve exact method with global optimal solution; however, CPLEX and mathematical model are not suitable for implementing a big data set. Algorithms and heuristics are required to improve the speed and satisfy the requirement of solving big problems.

5.3. Recommendation

All the suppliers, warehouses and customers are located in the Mekong Delta and only the factory is in Dong Nai Province. This location problem also contributed to the high transportation cost among pangasius feed. A suggestion for the company to install a factory specified in manufacturing aquaculture in Dong Thap, which is the major customer. The closer the factory to the customer is, the lower the transportation cost.

References

Altiparmak, F., Gen, M., Lin, L., & Paksoy, T. (2006). A genetic algorithm approach for multi-objective optimization of supply chain networks. *Computers & industrial engineering*, *51*(1), 196-215.

Anand Jayakumar, A., & Krishnaraj, C. (2015). Solving Supply Chain Network Gravity Location Model Using LINGO. *International Journal of Innovative Science Engineering and Technology*, *2*(4), 32-35.

Ballou, R. H. (2006). The evolution and future of logistics and supply chain management. *European Business Review*, 19(4), 332 – 348.

Boonmee, C., & Kasemset, C. Optimization of Distribution Network with Backhaul.

Dzupire, N. C., & Nkansah-Gyekye, Y. (2014). A multi-stage supply chain network optimization using genetic algorithms.

Fan Wang, Xiaofan Lai, Ning Shi, "*A multi-objective optimization for green supply chain network design*", p.1, 2011

Kocaoglu, Y., Cakmak, E., Kocaoglu, B., & Taskin Gumus, A. (2020). A Novel Approach for Optimizing the Supply Chain: A Heuristic-Based Hybrid Algorithm. *Mathematical Problems in Engineering*, *2020*.

Musa, R., Arnaout, J. P., & Jung, H. (2010). Ant colony optimization algorithm to solve for the transportation problem of cross-docking network. *Computers & Industrial Engineering*, *59*(1), 85-92.

Osman ALP, "An Efficient Genetic Algorithm for the p-Median Problem", Annals of Operations Research 122, 21-42, 2003.

R. Jamshidi, S.M.T. Fatemi Ghomi ∗, B. Karimi, "*Multi-objective green supply chain optimization with a new hybrid memetic algorithm using the Taguchi method*", p. 1, 2012

Tseng, Y. (2005). The role of transportation in logistics chain. *Eastern Asia Society for Transportation Studies*, 5, 1657–1672.

Wei-Chang Yeh, (2005). A hybrid heuristic algorithm for the multistage supply chain network problem. *The International Journal of Advanced Manufacturing Technology*, *26*(5), 675-685.

OPTIMAL STORAGE ALLOCATION FOR PRODUCTS IN WAREHOUSE A CASE STUDY OF YCH PROTRADE COMPANY FOR SUNTORY PEPSICO'S PRODUCT

Nguyen Le Binh Minh[1], ***Duong Vo Nhi Anh**[2], ***Ngo Thi Thao Uyen**[3]

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: [1]minhnguyenle2511 @gmail.com, [2]dvnanh@hcmiu.edu.vn,[3]nttuyen@hcmiu.edu.vn

Abstract: The warehouse plays a very important role in the logistics process of the production environment as well as in the service. Therefore, researchers have constantly developed problems of storage allocation for product in the warehouse in order to reduce use of human and make warehouse operation efficiency. YCH Protrade's warehouse is a unit specialized in providing logistics services, storing products in the warehouse produced by Suntory PepsiCo Viet Nam. YCH Protrade warehouse is having the problem of improper product locations and large travel distance of order picking process. The randomly arrange the product leads to wasting time searching for products, large travel distances and getting wrong item codes when preparing orders in warehouse, not finding the correct items in the location. Therefore, the purpose of the study is optimal storage allocation for products in warehouse based on the following factors: travel distance of order picking process, travel time and operation management in warehouse. Through learning about research articles on product storage policy in warehouse. I choose the dedicated storage policy according to the Storage Location Assignment Problem (SLAP) based on the ratio of import/export of each type of product in warehouse and ABC class - based storage policy based on the demand of the customer. The optimal storage policy helps to improve efficiency in order picking, avoid errors in product code selection, and especially reduces the travel distance of order picking process. Besides, the study will apply Travelling Salesman Problem (TSP) model to find the optimal route and minimize the total travel distance of two storage policy are: dedicated storage and ABC class - based storage. The IBM Cplex and Matlab will be used to build the model and solve this problem. Finally, the study uses the analytic hierarchy process (AHP) method based on the evaluation criteria which the results of the total travel distance and sub-criteria such as: travel time and operation management to choose the optimal product storage policy in warehouse.

Key words: dedicated storage policy, ABC class - based storage policy, SLAP model, Travelling Salesman Problem (TSP) model, AHP method.

1. Introduction

1.1. Background Overview

According to preliminary assessments, the capacity to design and operate an efficient warehouse in Vietnam is still very limited, leading to very high storage costs and wasted warehouse capacity. The most intuitive view of an inefficient warehouse is that goods are arranged messy, lack of storage area, damaged goods are stored in warehouse, large travel distance to prepare orders. The solution to improve the service level for the warehouse, increase the ability to strictly manage and make the most of the warehouse's resources to minimize storage costs is necessary. One of the factors affecting the operational efficiency of the warehouse is allocation the product storage location, the product is arranged according to the experience, causing a lot of loading time. According to Hall (2008) has applied different warehouse layout options based on the frequency of import and export to reduce the total travel distance for feasible results and high quantitative when applied in practice. The proposed improvement space could reduce the travel distance to pick up in stock and include the warehouse's scalability when designing.

1.2. Problem Statement

According to the monthly Key Performance Index (KPI) report, evaluate the internal operation of the warehouse has often been below regulatory levels in recent months, such as 91.17%/99.8% in 12/2020 or 93,88%/99.8% in 01/2021. The chart below shows the subjective and objective causes for the inefficient warehouse operations. In particular, there are two main reasons leading to this situation: improper product storage location and travel distance of order picking process in warehouse.

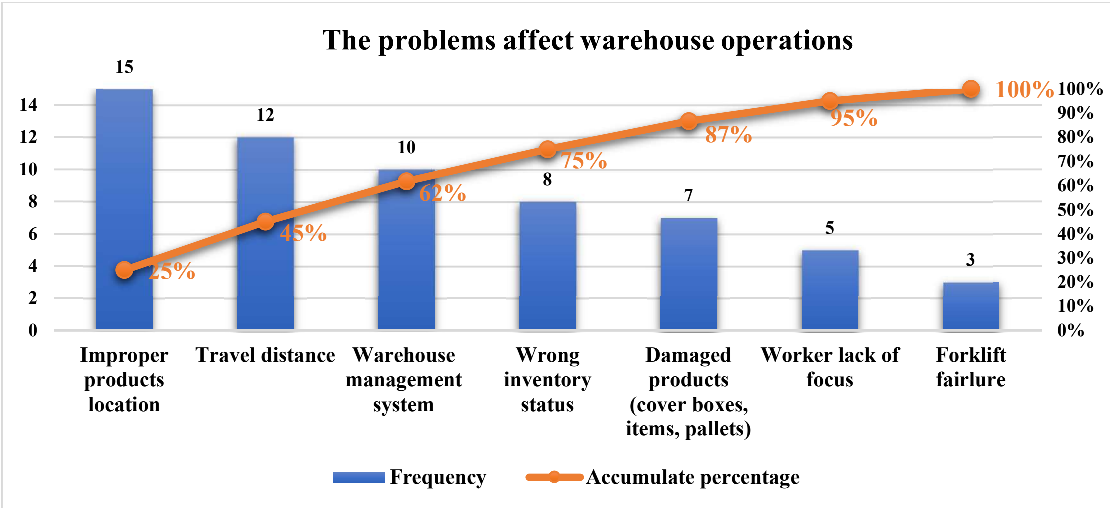


Figure 1: The Pareto Chart of problems affect warehouse operations

Firstly, improper products location is the most serious and frequent problem. SKUs are not coordinated and handled properly; the products are not arranged according to each separate item. They randomly arrange the products, if they see an empty position, then put it in and update the position on the system without giving a specific storage location for the product.

The second problem is the travel distance of order picking process affects warehouse operations. During the order picking process, picker goes to each picking locations according to the customer's order list. This causes waste distance during the total travel distance to pick up products in a day.

1.3. Objectives

The main objective of this study finds optimal storage allocation policy for products and minimize the travel distance in warehouse. Achieving this goal, the following steps should be taken.

The first thing about this study is to do the product storage policy in the warehouse. The storage allocation policy based on literature includes dedicated storage policy, and ABC class-based storage policy. The design option still has to meet the limitations of arranging enough goods according to demand, using only the right number of storage locations according to capacity and ensuring the aisle area in the warehouse.

The second thing is to minimize the total travel distance when order picking process by grouping order into batches in a single picking trip is applied.

After that used AHP method to choose optimal storage policy for products in warehouse. This based on criteria such as the travel distance by order picking process, sub-criteria: travel time and operation management.

1.4. Scope

The scope of this project is YCH Protrade's warehouse layout design, specifically in zone A in the warehouse. Products stored in the warehouse are beverage manufactured by Suntory PepsiCo's company.

The study only providing items delivered to small distribution channels, the limited product categories include: 11 beverages of Suntory PepsiCo's products equivalent to 50 SKUs.

2. Related Works

2.1. Literature Review

2.1.1. Dedicated Storage Policy

According to Muharni and Khoirunnisa (2019) has developed a dedicated storage policy involved in a steel manufacturing business. The issue that emerges in the warehouse of slab raw materials and in the placement of slab raw materials has no fixed guidelines. Dedicated storage will be used in the layout design of a slab raw material warehouse in this study to achieve the best layout and reduce material handling costs. The improvements in the raw material warehouse layout were introduced by considering the order of operations, class formation, actual warehouse position, and calculating the total distance of the movement. According to the research findings, in comparison to the current layout, it gives the best layout result with the least material handling cost.

2.1.2. Storage Location Assignment Problem Model (SLAP)

In this study of Ren-Qian Zhang (2019) is to use an established mathematical model strategy in the warehouse to assign storage location. The problem of storage location is solved with minimize warehouse travelling using integer programming. The storage location assignment problem (SLAP) involves determining the allocation of items to storage locations to maximize the order-picking efficiency. The SLAP assists in minimizing material handling and maximizing the use of space (De Koster et al., 2007, Gu et al.,2007). Storage location assignment is used to better allocate warehouse space to goods for familiarity and maximize the usage of space inside a warehouse.

Besides, the problem of Edmundo Salazar, José Luis Martínez (2017) in storage location allocation for a level 1 supplier with a Just In Sequence production method is the subject of this paper. The problem is characterized as finding the optimal item allocation at the lowest operating cost using a linear programming formulation. The ratio of the storage space needed by the demand frequency, also known as the cube-per-order-index, is a common attribute for assigning an element to a storage position. Correlated stock allocation, in which items that are often ordered together are mapped to locations similar to each other, is one approach that can help minimize travel distances (Calzavara, Glock, Grosse, Persona, & Sgarbossa, 2016). As a result, SLAP focuses on minimizing the overall frequency-weighted distance between the element storage areas and the picking area or production line, subject to similarity restrictions and constraints depending on individual situations (Xie, Mei, Ernst, Li, & Song, 2014).

2.1.3. ABC-Class Storage Assignment

According to Ruud H. Teunter; M. Zied Babai; Aris A. Syntetos (2010), ABC inventory categories are commonly employed, with demand value and volume being the most popular ranking criterion. In ABC applications, the normal procedure is to specify the same service level for all stock keeping units (SKUs) in a class. In this study, they show that using both demand value and demand volume as ABC ranking criteria, with constant service levels per class, leads to solutions that are far from cost optimum for three large real-world datasets. By giving the same service level to each SKU in a class, ABC categorization may be used to set service levels. This is consistent with findings from NONSTOP solutions (a company that specializes in demand chain optimization) and Pflitsch (2008) from SLIMSTOCK (a provider of forecasting and inventory management software, including inventory categorization software called ‘‘Slimstock ABC”). Both agree that the typical strategy is to fix service levels per class, based on their significant experience designing inventory control software.

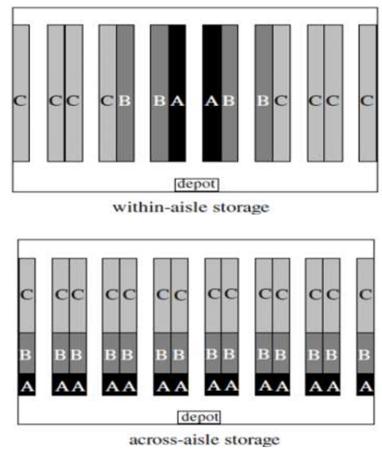


Figure 2: Illustration of two popular implementations of ABC class-based storage

2.1.4. Travelling Salesman Problem Model (TSP)

In this study of the objective of Sawik, T. (2016), the Travelling Salesman Problem mathematical model will be applied to find the optimal route or optimal travel distance for each order from customers. The asymmetric traveling salesman problem (ATSP) is a directed network issue in which only one direction of movement is allowed. A salesperson who starts his journey from node 0, the depot, must visit each node precisely once before returning to node 0. The challenge is to identify the shortest directed tour that visits n nodes in the smallest amount of distance.

2.1.5. Analytical Hierarchy Process (AHP)

The Analytical Hierarchy Process (AHP) of Kamal M. Al-Subhi, Al-Harbi (2001) is presented in this study as a viable decision-making tool for project management. The problem of contractor pre -qualification is provided as an example. The qualification criteria and contractors desiring to prequalify for a project are organized into a hierarchical framework. The approval criterion may be prioritized using the AHP, and a descending order list of contractors may be created to identify the best contractors for the job. To test the sensitivity of final conclusions to modest changes in judgments, a sensitivity analysis can be undertaken. The AHP is used to make group decisions in this study.

Using the 'Expert Choice' professional software, which is available commercially and developed for implementing AHP, the AHP implementation processes will be simplified. It is intended that this would encourage project management experts to use the AHP.

2.2. Conceptual Design Description

2.2.1. Dedicated Storage Policy

This is called a fixed location warehouse that includes the division of individual warehouse locations or the stocking addresses of each product. Since a warehouse location is divided or reserved for a particular product, it is called a dedicated warehouse. One of the advantages of dedicated storage is the data handling efficiency due to the fixed addressing of storage items. In order to minimize the total expected distance travelled the following approach is taken.

When a dedicated storage system is used, the number of positions assigned to a particular product should satisfy the maximum storage space requirement for that product type. In case of dedicated storage of many products, the required storage space is equal to the total maximum demand of all products. It is our job to define the storage space for all types of products in the warehouse. The maximum number of storage units indicates the required storage space for each product type. The storage space for all product types will be equal to the total required storage space.

2.2.2. Storage Location Assignment Problem (SLAP)

The SLAP concerns the allocation of products into a dedicated storage space and optimization of the material handling costs or storage space utilization. The goal is to find the best specific location for each item in the warehouse to store the products, thus minimizing the travel distance in the warehouse. The main optimization approaches concern warehouse space utilization and the cycle time for order preparation and picking operations, considering restrictions such as available storage capacity, order-picking resource capacities, and dispatching policies.

The SLAP as follows:

1. Information on the storage area, including its physical configuration and storage layout.
2. Information on the storage locations, including their availability, physical dimensions, and location.
3. Information on the set of items to be stored, including their physical dimensions, demand, quantity.

2.2.3. ABC Class - Based Storage Policy

This system groups items together and assigns them to storage locations. Both objects and storage areas are divided into the same number of classes. The level of turnover determines the item groups (like pick lines per time unit, or product units picked per time unit). In ABC-class, SKUs are classified as A (the small percentage of SKUs that account for the majority of activity), B (moderately important), or C (the bulk of the SKUs but only a small portion of the activity). The item classes are sorted according to decreasing turnover frequency, while the storage position classes are sorted according to increasing travel distance from the I/O stage. Following that, in this order, the item classes are allocated to the storage location classes (which should be large enough to accommodate the SKUs). Objects are randomly stored within a storage class. The main distinction between this approach and the volume-based assignment method is that this one assign objects to storage locations in groups. The volume-based and randomized storage assignment methods are combined in this process. Pareto's approach is a traditional method for categorizing objects in inventory management based on popularity.

2.2.4. Travel Distance Of Order Picking Process

The objective of travelling salesman problem algorithm is minimize total travel distance, therefore may be routing strategy also optimal policy. The travelling salesman problem is best way to find out the minimize of total travel distance of order picking process in of two storage policy: dedicated storage and ABC - class based storage by total travel distance of order picker within one month.

2.2.5. Analytical Hierarchy Process (AHP)

AHP is a multi-criteria decision making (MCDM) method proposed by Saaty (1980). It is a popular and widely used method for MCDM. Allows the use of qualitative, as well as quantitative criteria in the assessment. Based on pairwise comparison, AHP can be described with 3 main principles: analysis, evaluation, and synthesis. AHP first analyzes a complex, multi-criteria problem according to the hierarchical structure shown in the figure.

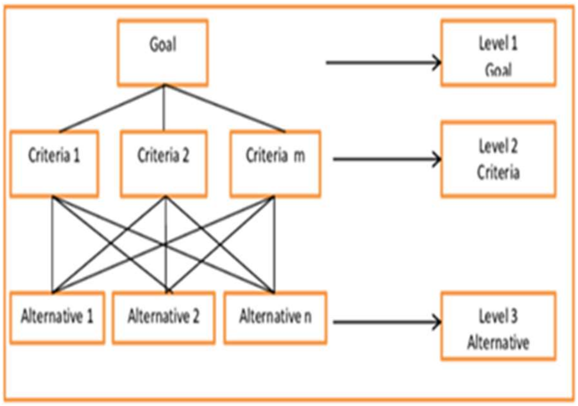


Figure 3: A Hierarchical Structure of AHP

A hierarchical structure diagram begins with the goal, which is analyzed through major criteria and component criteria, the final level usually includes possible alternatives. The evaluation process uses a pairwise comparison matrix with a 9-point scale, determines the weights based on the eigenvectors corresponding to the largest eigenvalues, and then checks the coefficient of consistency. Finally, all the weights are summed up to make the best decision. The process of analyzing and determining weights according to AHP is described in detail in the references (Saaty, 1980; Saaty et al. Vargas, 1994).

The study would like to introduce in more detail the application of AHP in choosing a product storage plan by providing criteria with objective evaluation rates according to the achieved data of the options in the research.

The AHP algorithm basically consists of two steps:

- Determine the relative weights of the decision criteria.
- Determine the relative rank (priority) of the alternatives.

The AHP consists of six fundamental phases, which are detailed below (Virendra Rajput, Dinesh Kumar, Arun Sharma, Shubham Singh, Rambhagat, 2018)

Step 1: AHP presents a complex choice problem as a series of minor subproblems. As a result, the first step is to divide the problem into a hierarchy, with a goal at the top, criteria and sub-criteria at levels and sublevels, and decision options at the bottom.

Step 2: Construct the choice matrix, which is based on Saaty's nine-point scale. The decision maker evaluates the priority score using Saaty's fundamental 1–9 scale. In this context, a score of 1 indicates equal significance, a score of 3 suggests considerably more significance, a score of 5 suggests substantially more important, a score of 7 shows significantly more significance, and a score of 9 indicates considerably more important. The numbers 2, 4, 6, and 8 are used to highlight important compromise values.

Table 1: The Fundamental scale for Pairwise Comparisions

The Fundamental Scale for Pairwise Comparisons		
Intensity of Importance	Definition	Explanation
1	Equal importance	Two elements contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favor one element over another
5	Strong importance	Experience and judgment strongly favor one element over another
7	Very strong importance	One element is favored very strongly over another; its dominance is demonstrated in practice
9	Extreme importance	The evidence favoring one element over another is of the highest possible order of affirmation
Intensities of 2, 4, 6, and 8 can be used to express intermediate values. Intensities 1.1, 1.2, 1.3, etc. can be used for elements that are very close in importance.		

Step 3: The third step is comparing the items of the built hierarchy in pairs. The goal is to determine their respective priority for each element at the next higher level. The pairwise comparison matrix is based on Saaty's 1–9 scale, with A denoting the pairwise comparison matrix, W being the eigenvector, and max being the highest eigenvalue. If items at higher levels of the hierarchy exist, the resultant weight vector is multiplied by the weight coefficients of those items until the top of the hierarchy is reached. The option with the greatest weight coefficient value should be considered the best. If n (n - 1)/2 comparisons are valid, where n is the number of criteria, then components aij will then meet the requirements aij = wi/wj = 1/aji and aii = 1 with I j, k = 1, 2, ..., n. The degree of preference of ith criterion over jth criterion is represented by aij in the comparison matrix. It appears that determining the weight of criteria through pairwise comparisons is more trustworthy than collecting them directly, because it is easier to make a comparison between two attributes than make an overall weight assignment.

Step 4: During the evaluation phase, AHP produces an inconsistency index (or consistency ratio) to indicate the consistency of decision makers' judgements. In both the decision matrix and pair - wise comparisons matrices, the inconsistency index may be determined using the equation: CI = (max – n) / (n – 1). The better the consistency, the nearer the inconsistency index is to zero. If the equivalence aij - aik = aik holds for all criteria, the evaluations will be consistent. To consider the AHP results as consistent, the appropriate index must be less than 0.10. If this is not the case, the decision-maker should return to Steps 2 and 3 and re-evaluate and compare the options again.

Table 2: Random indices

1	2	3	4	5	6	7	8	9	10
0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

Step 5: The comparison matrix must be normalized before all of the vector of priority computations can begin. As a result, each column must be divided by the total of the items in the next column. As a result, a normalized matrix with the total of the elements of each column vector equals 1 is generated.

Step 6: The eigenvalues of this matrix must be determined in order to determine the relative weights of criteria in the next section. This is a popular method in mathematics. The relative weights obtained in the third step should verify A.W = max. W

3. Methodology

3.1. Approach Comparison and Selection

Table 3: The summarize of advantages and disadvantages of approaches

No	Method	Pros	Cons
1	Dedicated Storage Policy	+ Order pickers become familiar with product locations → save time for pickers. + Typically reduces the material handling costs.	+ Must leave space for products not in warehouse. + It requires more storage space than random storage assignment in classes. + Must size locations for maximum warehouse.
2	Storage Location Assignment Model	+ The best specific location for each item. + Minimize travel distance. +Easy to establish and maintain. + Reduction of material handling and the improvement of space utilization. + Efficiency since can be customized and altered to the needs of one particular company.	+ This model cannot be applied if products are random and mixed storage policy
3	ABC-class storage assignment	+ Allocate products more efficiently in cycles and on scheduled dates. + The ABC class-based storage policy reduces the average one-way travel distance. + Leads to shorter travel time for storing and retrieving items.	+ Class-based storage requires more location than randomized storage. + Takes a long time to sort data into ABC class-based on the following factors: value, frequency,…
4	Asymmetric Traveling Salesman Problem (TSP)	+ Effective for the problem with a small number of nodes, TSP can be solved by exhaustive solution. + When compared to other global optimization approaches such as neural networks, genetic algorithms, and simulated annealing, TPS outperforms them. + It is possible to utilize it in dynamic applications (adapts to changes such as new distances, etc.)	+ TSP is computationally challenged to solve exponential time to convergence for a large number of nodes. + Convergence is definite, but the time it takes to reach it is unknown.
5	Analytical Hierarchy Process (AHP)	+ It allows multi criteria decision making. + It is applicable when it is difficult formulate criteria evaluations, i.e.:	+ Hierarchy is not always strict as should to be. + Interrelations between factors not flexible.

No	Method	Pros	Cons
		allows qualitative evaluation as well as quantitative evaluation. + Consistency in evaluation. + Straightforward and convenient. + Simplicity by using pairwise comparisons. + Versatility, independence, measurement.	+ Conflict between decision makers. + Decision maker capacity. + Many qualitative components are not convincing difficult to determine the weight when there are too many indicators.

3.2. Research Process

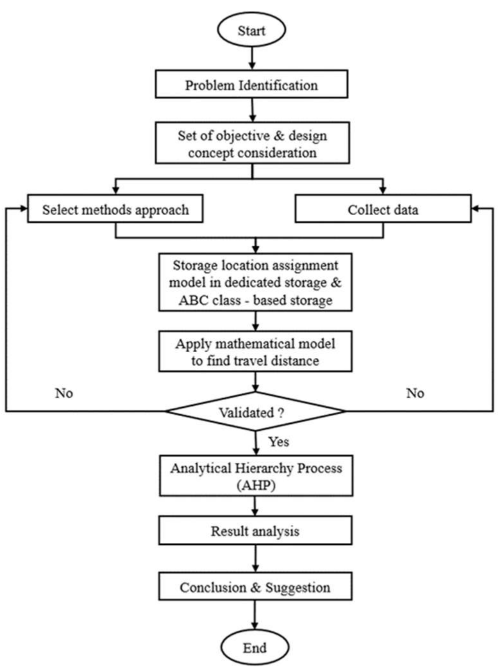


Figure 4: The framework planned of the research

Problem Identification: Firstly, observe the current state of the warehouse, what problems are encountered and identify the main problem of the warehouse. From understanding the state of the warehouse, applying the pareto chart to find out the main causes affecting the warehouse operation. Secondly, propose methods to approach and solve the goal of this problem.

Set of objective and design concept consideration: After determining the problem statement, proceed to determine the object of the problem. Then, find specific methods to be able to solve the problem of the research. The problems explored are related to the assignment of product locations and minimize travel distance in the warehouse are essential. All options will be compared and choosing the most suitable method to solve the problem. Finally, finding the ultimate optimal solution to the problem.

Select model approach: There are many methods to solve the problem, but we should consider choosing the best solution and most convenient method. For the research papers found, I will choose 2 storage assignments to conduct: ABC class - based storage and storage location assignment model for the allocation of a dedicated warehouse. Nextly, apply the algorithm model to find minimize the total travel distance of order picking process of two product storages in warehouse.

Collect data: After identifying the model approach, we proceeded to collect the necessary data about the current warehouse to solve the problem. Particularly: the number of positions in the storage need, the number of products, the SKU, the rate of import/export of products, the distances in warehouse, the size of products, pallets, storage area, aisle area, warehouse layout, amount of storage, the orders picking of customer, …

Storage location assignment model in dedicated storage & ABC class – based storage

• **Storage location assignment model in dedicated warehouse:** This study will apply the input/output ratio algorithm to the warehouse layout based on James A. Tompkins (2010). It is an algorithm where the products are arranged in the input/output positions in the warehouse to minimize the execution time in import/export products. There should be enough warehouse locations to storage all products. The objective is to minimize travel distance when import in warehouse as well as when export from warehouse for products in warehouse, the problem uses the advantage of controlling the large volume of a product, a product with a high frequency. If products with high input/output frequency are assigned in location with short distance, it will reduce the total travel distance of order picking process.

• **ABC-class based storage:** ABC analysis is used to perform classification for products on locations in the warehouse. The products are prioritized in order to arrange them in good positions for them in warehouse locations based on demand of customer. When arranging must meet enough demand for a new product class to move to another product of another class. In addition, in order to make good use of storage for classification, it is necessary to consider priority points for convenient import and export to arrange products.

Apply mathematical model to find travel distance of order picking process: Use a mathematical model to minimize the total travel distance of order picking process in the 2 types of storage found above: dedicated storage and ABC-class based storage.

Validated?: If the problem achieves suitable results, then conduct an analytical hierarchy process (AHP) methods and evaluate. If not, review the data information or the methods approach.

Analytical Hierarchy Process (AHP): The AHP method has been widely applied to many fields such as natural, economic, social, medical, etc. It is used as a flexible tool for decision analysis with many criteria, allowing clearly see the evaluation criteria and decide many attributes, which refers to a quantitative technique. The hierarchical analysis process can consider multiple sub-criteria simultaneously with groups of criteria and can combine both qualitative and quantitative analysis. Based on the principle of pair comparison, the AHP method can be described with three main principles: analysis, evaluation, and synthesis. AHP answers questions like “Which option should we choose?” or “Which option is best?” by selecting the best alternative that satisfies the decision maker's criteria on the basis of comparing pairs of alternatives and a specific computational mechanism. Therefore, I will use the AHP method based on the evaluation criteria are travel distance and sub-criteria to choose the optimal product storage policy in warehouse.

Result Analysis: This section will present the results and analysis of the travel distance of order picking process of the 2 types of product storage. Finally, the results choose the optimal product storage policy based on the evaluation criteria of AHP methods. Besides, evaluate how the results affect the economic and social aspects (labor, insurance, safety, ...)

Conclusion & Suggestion: On the basis of the analysis of the above results, the results will be evaluated and proposed options that the warehouse should consider applying to achieve efficiency as well as suggest more problems that the warehouse is facing that needs to be researched to solve.

4. Implement The Plan And Modelling The Problem

4.1. Data Collection

4.1.1. The Storage Required Of Each Item Category

Another required parameter when modeling the warehouse reassign storage locations problem is the amount of storage demand. From there, determine the number of slots needed for each product. The storage requirement of an item is understood as the number of slots needed in a warehouse to store all products of that item category. Therefore, different product categories will require different number of requirements.

After collecting data from the company, we have the quantity types of product storage in warehouse from August 2020 to December 2020.

Table 4: The quantity types of product storage in warehouse from August 2020 to December

No	Name of item	Unit	Aug	Sep	Oct	Nov	Dec	Maximum storage needs
1	Pepsi Cola	Carton	6180	6240	6120	6300	6480	6480
2	7-Up Standard	Carton	5520	5460	5700	5400	5760	5760
3	Mirinda Orange	Carton	5340	5280	5400	5460	5520	5520
4	Evervess Soda	Carton	1800	1980	1860	1920	2040	2040
5	Sting Standard	Carton	5520	5700	5580	5640	5760	5760
6	Lipton Ice Tea	Carton	3900	3840	4020	3960	4080	4080
7	CC Lemon	Carton	1800	1920	1860	1980	2160	2160
8	O-Long Tea Plus	Carton	1800	1680	1860	1740	1920	1920
9	Twister Orange	Carton	1700	1600	1550	1650	1800	1800
10	Aquafina Standard	Carton	3936	3888	3840	3984	4032	4032
11	Mountain Dew	Carton	1240	1320	1360	1400	1440	1440
Pallet demand			678	681	685	691	718	

From the need for quantity and storage specifications, the maximum storage requirement (number of pallets) in highest month for each product is calculated.

Table 5: The storage requirement of products according to maximum storage need

No	Name of item	Maximum storage needs	Total number of cartons per pallet	Storage requirement (slots)
1	Pepsi Cola	6480	60	108
2	7-Up Standard	5760	60	96
3	Mirinda Orange	5520	60	92
4	Evervess Soda	2040	60	34
5	Sting Standard	5760	60	96
6	Lipton Ice Tea	4080	60	68
7	CC Lemon	2160	60	36
8	O-Long Tea Plus	1920	60	32
9	Twister Orange	1800	50	36
10	Aquafina Standard	4032	48	84
11	Mountain Dew	1440	40	36
Total				718

From the above analysis table, the storage requirement of the warehouse is 720 pallets which is equivalent to 720 locations. In particular, the maximum storage needs of 718 pallets. Therefore, the storage capacity of the warehouse still meets the maximum storage needs.

4.1.2. The Throughput of Each Item Category

The input and output of each item category over a period of time is referred to as throughput. Based on the collected data, it can be seen that the throughput of each item category changes in different time periods. The throughput rates of Sting Standard, Aquafina Standard, Pepsi Cola, and 7-Up Standard are higher than other Suntory PepsiCo products.

Table 6: The throughput of each item category (some typical items)

Name of item	Item code	Throughput (items/ month)	Item description	Item code	Throughput (items/ month)
Sting Bottle 24x240 ml	S-1101	73	7-Up Standard Bottle 24x207 ml	U-1401	54
Aquafina Standard Pet 24x355 ml	A-1201	65	O-Long Tea Plus Pet 24x455ml	O-1501	58
Pepsi Cola Bottle 24x207 ml	P-1301	66	Lipton Ice Tea Can 24x330 ml	L-1601	12
Mirinda Orange Bottle 24x207 ml	M-1701	22	Twister Orange Can 24x320 ml	T-1901	10
CC Lemon Can 24x330ml	C-1801	20	Mountain Dew Pet 24x500ml	D-1001	18
Evervess Soda Bottle 24x300 ml	E-2001	8	Evervess Soda Can 24x330 ml	E-2002	6

4.1.3. The Percentage of Travel In/Out Between Docks And Warehouse

The percentage of travel in/out between docks and warehouses. On the other hand, is a particular way of contributing of one dock to the total throughput speed in warehouse. Therefore, to assess the feasible of the model, current scenario for traveling between docks and locations are attempted.

Table 7: Percentage of travel in/out between dock and warehouse in current state

Dock	In/Out
Percentage	100% = 1

4.1.4. Distance Between Docks And Storage Locations

Every SKU is assigned a unique and specific location within warehouse. The first is to calculate the distance matrix for each location in the warehouse. Currently, the warehouse has one import/export dock.

Due to the limitation of running points in Cplex, 4 positions will be merged into one group. After grouping 4 positions into one group, instead of 720 positions, there will be 180 positions. The analysis area includes 180 locations and 6 aisles. The 4 aisles in the middle of the layout, pickers can picking on 2 sides.

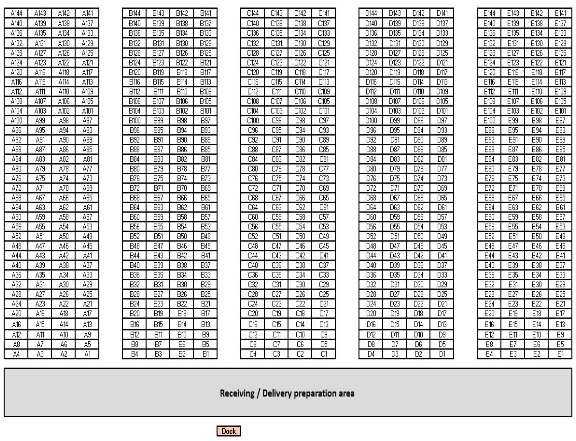


Figure 5: Specific storage locations of SPVP's product in warehouse

4.2. Types Of Product Storage Policies

4.2.1. The Dedicated Storage Policy

4.2.1.1. Assumtion And Notation

The dedicated strategy assigns each item to a dedicated storage location according to its individual properties, such as popularity, turnover, volume, or cube-per-order index (Petersen et al., 2005). Apply the storage location assignment problem (SLAP) involves determining the allocation of items to storage locations and maximize the order-picking efficiency of dedicated storage in the warehouse.

Each empty storage location is equally usable when the storage operation is performed. The movement of the forklift in the warehouse is assumed to be linear motion.

In this section, the notation that is used in the model, which are:

- n = the number of storage locations
- o = the number of items
- s = the number of input/output points (docks)
- S_i = the number of storage locations required for items i
- T_i = number of trips in/out (throughput) of storage for item i
- p_j = percentage of travel in/out of storage locations to/from I/O points j
- d_{jm} = distance required between storage location m and docks j
- x_{im} = 1 if item i is assigned to storage location m; otherwise, 0

4.2.1.2. Mathematical Model

The storage location assignment model is based on the 2010 Facility Planning 4th edition of Jame A. Topmpkins:

$$\text{Minimize } \sum_{i=1}^{o} \sum_{m=1}^{n} \left(\frac{Ti}{Si}\right) \sum_{j=1}^{s} p_j d_{jm} x_{im}$$

Subject to:

- Only one item can be stored in each storage location
- $\sum_{i=1}^{o} x_{im} = 1$ for m = 1, …, n
- After the assigning procedure, the total number of storage locations for item i must be equal to the initial requirements number of storage locations for item i.
- $\sum_{m=1}^{n} x_{im} = S_i$ for i = 1, …, o
- x_{im} = (0,1) for all i and m

4.2.1.3. The Layout Of Product Locations In Dedicated Storage

CPLEX Optimization Studio is used to write mathematical models in this case. The problem is defined with mathematical model to find the optimal allocation of items at a minimize travel distance. The result will be that each SKU will be assigned to each specific storage location in the warehouse.

Based on the code results, we get the items locations priority order from high to low. The items are arranged at the location of the highest Tj/Sj ratio to the location with the smallest distance, when the pallet demand of a product is fully met, then it is transferred to the ordered item lower priority.

Due to the limitation of running points in Cplex, 4 positions will be merged into 1 group (representing 1 position in Cplex). After grouping 4 positions into 1 group, there will be 180 positions instead of the original 720 positions.

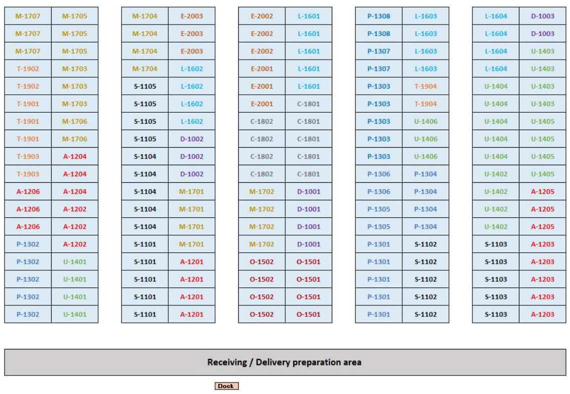


Figure 6: The storage location of items in dedicated storage in the warehouse

4.2.2. The ABC Class – Based Storage Policy

ABC class - based storage policy assigns the most frequently requested SKUs to the best locations on the rack face. ABC class - based storage policy is simple and easy to apply. Items are classified into classes based on how many times they appear in an ordered set. Class A items are relatively small in quantity but account for a large amount of the activity, while Class C items are relatively large but account for a relatively small amount of the activity. Items between the two upper classes form class B. Specifically in the groups are:

- Group A accounts for 20% of total product volume in warehouses but about 50% of demand of customer.
- Group B accounts for 30% of total product volume in warehouses but about 30% of demand of customer.
- Group C accounts for 50% of total product volume in warehouses but about 20% of demand of customer.

Based on customer demand in December 2020, to divide SKUs into classes A, B and C based on ABC class - based storage policy.

Table 8: ABC classification from the demand of customer in December 2020

No	SKU's Code	Demand	% Of Total Demand	% Cumulative	ABC Classify
1	S-1101	1020	8.44	8.44	A
2	S-1102	818	6.77	15.20	A
3	O-1501	578	4.78	19.99	A
4	O-1502	544	4.50	24.49	A
5	P-1301	768	6.35	30.84	A
6	P-1302	681	5.63	36.47	A
7	A-1201	795	6.58	43.05	A
8	U-1401	845	6.99	50.04	A
9	S-1103	611	5.05	55.09	B
10	S-1104	243	2.01	57.10	B
11	P-1303	297	2.46	59.56	B
12	P-1304	248	2.05	61.61	B
13	A-1203	556	4.60	66.21	B
14	M-1701	303	2.51	68.72	B
15	M-1702	276	2.28	71.00	B
16	U-1402	222	1.84	72.83	B
17	U-1404	169	1.40	74.23	B
18	U-1405	126	1.04	75.28	B
19	D-1001	302	2.50	77.77	B
20	C-1801	314	2.60	80.37	B
21	S-1105	72	0.60	80.97	C
22	A-1202	115	0.95	81.92	C
23	A-1204	76	0.63	82.55	C
24	A-1205	154	1.27	83.82	C
25	A-1206	144	1.19	85.01	C
26	P-1305	101	0.84	85.85	C
27	P-1306	125	1.03	86.88	C
28	P-1307	76	0.63	87.51	C
29	P-1308	72	0.60	88.10	C
30	U-1403	104	0.86	88.97	C
31	U-1406	73	0.60	89.57	C
32	L-1601	121	1.00	90.57	C
33	L-1602	69	0.57	91.14	C
34	L-1603	69	0.57	91.71	C
35	L-1604	78	0.65	92.36	C
36	M-1703	54	0.45	92.80	C
37	M-1704	52	0.43	93.23	C
38	M-1705	72	0.60	93.83	C
39	M-1706	74	0.61	94.44	C
40	M-1707	50	0.41	94.85	C
41	C-1802	100	0.83	95.68	C
42	T-1901	65	0.54	96.22	C
43	T-1902	50	0.41	96.63	C
44	T-1903	63	0.52	97.15	C
45	T-1904	50	0.41	97.57	C
46	D-1002	102	0.84	98.41	C
47	D-1003	42	0.35	98.76	C
48	E-2001	38	0.31	99.07	C
49	E-2002	64	0.53	99.60	C
50	E-2003	48	0.40	100	C

4.2.2.1. The Layout Of Product Locations In ABC Class - Based Storage

Based on the demand of customer, assign products to specific locations in the warehouse. Class A will include the SKU with the highest demand and are located near the dock. Class B includes the SKU with the second highest commodity demand. Finally, class C has the lowest demand.

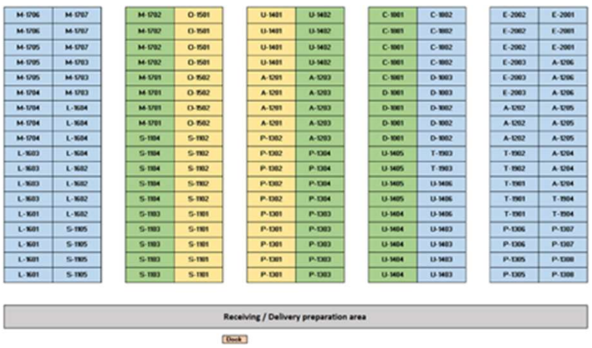


Figure 7: The storage location of items in ABC class - based storage in the warehouse

4.3. Travel Distance of Order Picking Process

4.3.1. Sets

- N: set of location includes pick location and dock (0: dock, n+1: dummy dock)
- P: set of products storages in warehouse.
- i: from location
- j: departure location

4.3.2. Parameter

- d_{ij}: distance from item's position i to item's position j
- c: capacity per picker (maximum = 70 cartons)
- o_p: = 1 if product p in picking order
- a_{pi}: = 1 if product p is placed in item's position i, otherwise is 0, (product p have many cartons in many places in warehouse, so the picker only needs to stop in any position which have product to pick)
- m_i: mass (quantity) of item in picking order at item's location i

4.3.3. Decision Variable

- x_{ij}: = 1 if the path from item's position i to item's position j, otherwise, 0
- y_i : = 1 if item's position i is have to visit, otherwise, 0
- s_i: current cartons of picker before going to item's position i

4.3.4. Objective Function

The purpose of this research is to analysis order picking efficient of two storage policy: dedicated storage and ABC - class based storage by total travel distance of order picker within one month. Therefore, the objective function of travelling salesman problem is best way to find out the minimize of total travel distance of order picking process.

$$Minimize\ Z = \sum_{i=0}^{n+1}\sum_{j=0}^{n+1} d_{ij} * x_{ij} \qquad (1)$$

4.3.5. Constraints

Constraint 1: Only pick what is assign

$$\sum_{\substack{j=1 \\ j\neq i}}^{n+1} x_{ij} = y_i \quad \forall i > 0, i \leq n, i \in N \qquad (2)$$

Constraint 2: Ensure flow in equal flow out

$$\sum_{\substack{i=0 \\ i\neq h}}^{n} x_{ih} - \sum_{\substack{j=1 \\ j\neq h}}^{n+1} x_{hj} = 0 \quad \forall h > 0, h \leq n, h \in N \qquad (3)$$

Constraint 3: Assign picker where to pick product p

$$\sum_{i=1}^{n} y_i * a_{pi} = o_p \quad \forall p \in P \qquad (4)$$

Constraint 4: Calculate the current boxes of picker before going to item's position i. The constraint also eliminates sub-tour

$$s_j \geq s_i + m_i * x_{ij} - c * (1 - x_{ij}) \quad \forall i, j \in N \qquad (5)$$

Constraint 5: Capacity constraint

$$m_i \leq s_i \leq c \quad \forall i \in N \qquad (6)$$

5. Results And Analysis

5.1. Result Travel Distance Of Order Picking Process

After running the model with Cplex software, the data solution is obtained. The goal of the model is to minimize the total distance traveled to pick up goods of two types of warehouses: ABC class-based storage and dedicated storage. After obtaining the travel distance of the two types of warehouses, the comparison with the original travel distance in random storage. The travel distance of the random storage is taken from the actual collection. The solutions of the 3 types of storage are shown below.

Table 9: The total travel distance of three types of product storage in the warehouse

	Random	ABC	Dedicated
Day 1	753.4	529.2	456.4
Day 2	780	546	305.2
Day 3	798	556	436.8
Day 4	750.6	484.4	394.8
Day 5	806.8	579.6	476
Day 6	757.6	512.4	420
Day 7	745.4	509.6	509.6
Day 8	784.6	551.6	464.8
Day 9	747.4	512.4	414.4
Day 10	784.6	562.8	400.4
Day 11	761.2	520.8	445.2
Day 12	776.4	551.6	403.2
Day 13	726.8	470.4	439.6
Day 14	801.4	571.2	392
Day 15	752	498.4	434
Day 16	757.6	481.6	375.2
Day 17	769.4	504	386.4
Day 18	761	506.8	434
Day 19	786	557.2	434
Day 20	745	518	400.4
Day 21	789.6	520.8	476
Day 22	759.6	492.8	448
Day 23	794.2	557.2	464.8
Day 24	754.4	529.2	336
Day 25	784.2	599.2	428.4
Day 26	762.6	537.6	490
Day 27	743.4	501.2	422.8
Day 28	814.4	509.6	389.2
Day 29	743.6	498.4	389.2
Day 30	764.6	548.8	445.2
Total	23055.8	15818.8	12712

Table 10: Percentage of saving the total travel distance

Type	Random	ABC	Saving	% Saving
Total Travel Distance	23055	15818	7237	31.3%

Type	Random	Dedicated	Saving	% Saving
Total Travel Distance	23055	12712	10343	44.8%

From the table above, we can see that the travel distance in the ABC class-based storage is 15818m, saving 31.3% compared to the original warehouse is random storage. Meanwhile, in the dedicated warehouse is 12712m, saving up to 44.8% of the travel distance compared to the original warehouse.

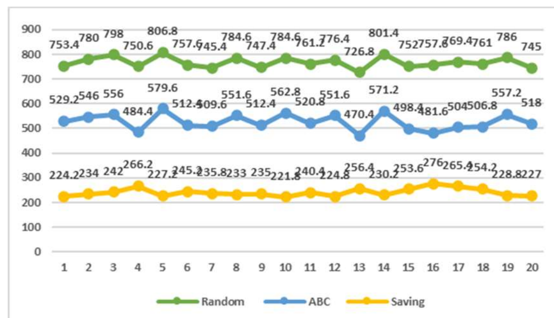


Figure 8: Comparison chart about travel distance each day of random storage and ABC class – based storage policy

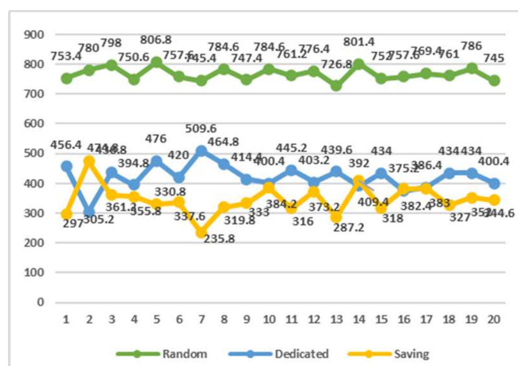


Figure 9: Comparison chart about travel distance each day of random storage and dedicated storage policy

The value total travel distance of the two storages: dedicated storage and ABC class – based storage compared to the original (random storage) shows:

- The total distance traveled by the ABC storage policy saves 31.3% of the travel distance compared to the original warehouse (random storage).
- Meanwhile, for the dedicated storage policy, it saves up to 44.8% of the travel distance compared to the original warehouse.

The best storage in terms of total travel distance and distance savings is dedicated storage policy.

- Instead of each order being selected individually, it saves a lot of effort when the orders are grouped together. With the application of this method to the YCH warehouse, it will save a lot of labor for workers.
- Besides, another advantage brought in minimizing travel distance is saving time. If the total travel distance decreases, the time to complete the order will also decrease.

5.2. Apply AHP Method to Get The Final Result

To choose the optimal storage policy, many factors need to be considered, so we use the AHP method for analysis. Among the factors, there are 3 important factors: travel distance, travel time and operation management in the warehouse.

- The first is to compare with the original storage, we can see that the initial goal of reducing the total travel distance in the warehouse has been effectively achieved, the dedicated storage is 12712m (saving 44.8%) and the ABC class – based storage class is 15818m (saving 31.3%) compared to the original total distance.
- The second is the travel time in the warehouse. If the total travel distance decreases, the travel time to complete the order also decreases.

Finally, in terms of operation management, according to the theory of dedicated storage, there are more advantages when the products locations are clear and fixed, workers can remember the location of products, without spending much time searching and the manager can do a monthly inventory. With an ABC class – based storage, it is impossible for workers to remember the correct product location as the locations change over time, they will have to spend time searching when they want to prepare an order. Usually, the ABC class – based inventory will require a lot of resources when the inventory and management will aggregate it to compare with the data in the computer.

Table 11: Determine priority for criteria

Criteria	Random storage	Dedicated storage	ABC class-based storage
Total travel distance	23055	12712	15818
Travel time	Longest travel time due to longest travel distance	Shortest travel time among storage types due to shortest travel distance.	The travel time of ABC storage is faster than that of random storage.
Operational management	Difficult because the product locations are assigned messy	Easy because the product position is fixed	Difficulty when the product position is random on a specified layer, the location changes over time, demand.

The study uses the AHP method to propose the option to choose the most suitable storage policy in the warehouse that meets the criteria. The criteria are calculated by giving rates based on comparison of quantitative data. In this thesis, using the criteria is the goal of the thesis and the data has been calculated above is total travel distance. Besides, with the sub-criteria of operation management, time helps to make more accurate decisions.

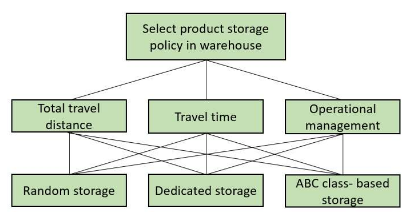


Figure 10: The criteria diagram and plan to apply the AHP method

Table 12: Evaluation of the criteria

	Total travel distance	Travel time	Operational management
Total travel distance	1	3	5
Travel time	$\frac{1}{3}$	1	3
Operational management	$\frac{1}{5}$	$\frac{1}{3}$	1
Sum (S)	$\frac{23}{15}$	$\frac{13}{3}$	9

Table 13: Priority vector of the criteria

	Total travel distance	Travel time	Operational management	Priority vector (PV)
Total travel distance	$\frac{15}{23}$	$\frac{9}{13}$	$\frac{5}{9}$	0.63
Travel time	$\frac{5}{23}$	$\frac{3}{13}$	$\frac{3}{9}$	0.26
Operational management	$\frac{3}{23}$	$\frac{1}{13}$	$\frac{1}{9}$	0.11

The Eigen value is:

Λmax = Sum x Priority vector = $\begin{bmatrix} \frac{23}{15} & \frac{13}{3} & 9 \end{bmatrix} \times \begin{bmatrix} 0.63 \\ 0.26 \\ 0.11 \end{bmatrix} =$ 3.08

The consistency index: $CI = \frac{\Lambda \max - n}{n-1} = \frac{3.08-3}{3-1} = 0.04$

The consistency ratio: $CR = \frac{CI}{RI} = \frac{0.04}{0.58} = 0.067 < 10\%$ (acceptable)

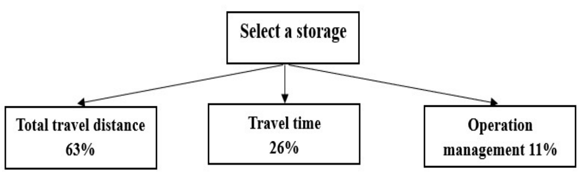


Figure 11: The chart of evaluation of the criteria

❖ **Pair wise comparation for alternatives:**
1. Total travel distance

The total travel distance will correspond to the following ratios:

➢ Dedicated storage compares with ABC storage: $\frac{12712}{15818} = \frac{4}{5}$

➢ Dedicated storage compares with random storage: $\frac{12712}{23055} = \frac{11}{20}$

➢ ABC storage compares with random storage:

$\frac{15818}{23055} = \frac{17}{25}$

Table 14: Evaluation of the alternatives based on the total travel distance

Total travel distance (1)	Random	Dedicated	ABC
Random	1	$\frac{11}{20}$	$\frac{17}{25}$
Dedicated	$\frac{20}{11}$	1	$\frac{5}{4}$
ABC	$\frac{25}{17}$	$\frac{4}{5}$	1
Sum (S)	$\frac{802}{187}$	$\frac{47}{20}$	$\frac{293}{100}$

Table 15: Priority vector of alternatives based on the total travel distance

Total travel distance (1)	Random	Dedicated	ABC	Priority vector (PV)
Random	$\frac{187}{802}$	$\frac{11}{47}$	$\frac{68}{293}$	0.23
Dedicated	$\frac{170}{401}$	$\frac{20}{47}$	$\frac{125}{293}$	0.43
ABC	$\frac{275}{802}$	$\frac{16}{47}$	$\frac{100}{293}$	0.34

The consistency index: $CI = \frac{\Lambda \max - n}{n-1} = \frac{2.993-3}{3-1} = -0.0035$

The consistency ratio: $CR = \frac{CI}{RI} = \frac{-0.0035}{0.58} = -0.006 < 10\%$ (acceptable)

2. Travel time

Table 16: Evaluation of the alternatives based on the travel time

Travel time (2)	Random	Dedicated	ABC
Random	1	$\frac{1}{6}$	$\frac{1}{4}$
Dedicated	6	1	$\frac{3}{2}$
ABC	4	$\frac{2}{3}$	1
Sum (S)	11	$\frac{11}{6}$	$\frac{11}{4}$

Table 17: Priority vector of alternatives based on the travel time

Travel time (2)	Random	Dedicated	ABC	Priority vector (PV)
Random	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{1}{11}$	0.09
Dedicated	$\frac{6}{11}$	$\frac{6}{11}$	$\frac{6}{11}$	0.55
ABC	$\frac{4}{11}$	$\frac{4}{11}$	$\frac{4}{11}$	0.36

The consistency index:

$$CI = \frac{\Lambda\max - n}{n-1} = \frac{2.988-3}{3-1} = -0.006$$

The consistency ratio:

$$CR = \frac{CI}{RI} = \frac{-0.006}{0.58} = -0.01 < 10\%$$

(acceptable)

3. Operation Management

Table 18: Evaluation of the alternatives based on the operation management

Operation Management (3)	Random	Dedicated	ABC
Random	1	$\frac{1}{5}$	$\frac{1}{3}$
Dedicated	5	1	$\frac{5}{3}$
ABC	3	$\frac{3}{5}$	1
Sum	9	$\frac{9}{5}$	3

Table 19: Priority vector of alternatives based on the operation management

Operation Management (3)	Random	Dedicated	ABC	Priority vector (PV)
Random	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0.11
Dedicated	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	0.56
ABC	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0.33

The consistency index:

$$CI = \frac{\Lambda\max - n}{n-1} = \frac{2.988}{2} = -0.006$$

The consistency ratio:

$$CR = \frac{CI}{RI} = \frac{-0.006}{0.58} = -0.1 < 10\%$$

(acceptable)

After all criteria matrices and consistency ratios are accepted, aggregated into the evaluation matrix, and multiplied by the evaluation weight matrix, we get the ranking of the options according to AHP below

Priority vector of alternatives x Priority vector of criteria = Weights

	(1)	(2)	(3)
Random	0.23	0.09	0.11
Dedicated	0.43	0.55	0.56
ABC	0.34	0.36	0.33

X

(1)	0.63
(2)	0.26
(3)	0.11

Table 19: Ranking of alternatives according to AHP

Types of storage	Weights	Rank
Random	0.18	3
Dedicated	0.47	1
ABC	0.35	2

Based on the AHP method, the study chooses the dedicated storage policy that is considered the most optimal in all types of storage. Specifically, consider 3 factors to choose the optimal storage policy: total travel distance, travel time and operation management in the warehouse.

5.3. Summary Of Results

As mentioned at the beginning of the topic, "Optimal Storage Allocation For Products In Warehouse" was formed according to the trend of improving operational management for warehouses. The topic is proposed to solve the difficult problem that is the situation of messy arrangement of goods, large travel distances, leading to reduced productivity in warehouses, which directly affects competitiveness of the company. In an effort to come up with an effective and practical solution, the product storage plans given based on logical calculations have completed the initial objectives. Below is a summary of what has been done and the results obtained from finding the optimal storage policy for the warehouse:

- Calculate the storage capacity and allocate product locations in the warehouse to meet the original design, ensuring the aisle in the warehouse.
- Minimize travel distance for product storage policies: random storage, dedicated storage, and ABC class - based storage.
- Proposing two product storage plans based on the frequency of in and out of each type of product to reduce the total travel distance: dedicated storage of 12712m (saving 44.8 %) and ABC class – based storage of 15818m (saving 31.3%) compare with the original total distance.
- Select an optimal storage plan by AHP method to choose the product storage policy in the warehouse. A storage policy is considered optimal based on many factors such as: short total travel distance, fast travel time, easy operation management in the warehouse, ... One of the factors decided it was the travel distance by the worker. Therefore, by analyzing the total travel distance for each storage type test to see which type of storage is more efficient.

The end result of this research is that dedicated storage policy is selected, contributing to the reduction of the travel distance compared to the original storage (random storage) by 44.8%.

6. Conclusion And Recommendation

6.1. Conclusion

The topic correctly identifies the main problem at the company, choosing the right object to deploy is the storage policy at the warehouse, the object that directly affects the productivity of the warehouse. Therefore, the research is assessed by the company as being able to be deployed. With the solution to optimize the product storage location, it has helped to reduce the total distance of moving workers, improve the efficiency of using the warehouse area, improve the warehouse operation capacity and step by step to improve KPI of warehouse.

The proposed product storage options are implemented on specific theoretical bases, according to the constrained conditions of the warehouse, and both options meet the original objectives. Compare each option according to specific evaluation indicators suitable to the objectives, properly assess the advantages and disadvantages of each option according to objective ratios. This ensures support for decision-making on the choice of implementation plan for the company's management.

The data collection encountered difficulties due to the privacy policy and lack of practical experience which greatly influenced the process of making my thesis.

However, I also achieved some results after more than 3 months of hard work:

- Understand the warehouse operation process, the problem in the warehouse that the company is facing.
- Expand your practical knowledge and improve your ability to work independently.
- Apply the knowledge learned to practical problems: calculate the distance traveled of order picking, apply the AHP method to evaluate different types of criteria, ... then choose the optimal storage policy for the warehouse.

6.2. Recommendation

In the operation management in the warehouse, there are many related issues that need to be resolved. However, due to the short implementation time of the thesis, the problems of the warehouse have not been completely solved, the thesis only solves the most necessary basic problems that directly affect the operation of the warehouse. Therefore, in the future, when applying the topic, it is necessary to research to develop, expand and supplement. Here are some directions for further development of the topic:

- Apply models to reduce inventory while ensuring order fulfillment.
- Forecasting customer needs to make a reasonable import/export plan, avoiding redundancy.
- Designing working operations during order preparation, reducing redundancies, and considering arising problems also helps to improve operational safety.
- Design an overall quality inspection process for each product type, quality assessment criteria, summarizing each type of defect for each product and how the defect identification characteristics can be classified by workers. correct errors and not too dependent on quality inspectors.

- Need to buy more protective gear or make regulations on clothes for workers when working at the warehouse because the goods in the warehouse are almost heavy and bulky.

Reference

Battini, Daria; Glock, Christoph H.; Grosse, Eric H.; Persona, Alessandro; Sgarbossa, Fabio (2016), "Human energy expenditure in order picking storage assignment: A bi-objective method" *Computers & Industrial Engineering*, S0360835216300134

Edmundo Salazar, José Luis Martínez (2017), "Storage location allocation model for a just in sequence (JIS) production system". *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Bogota, Colombia,

Jing Xie, Yi Mei, Andreas Tilman Ernst, Xiaodong Li, Andy Song (2015), "Scaling up solutions to storage location assignment problems by genetic programming"

Kamal M. Al-Subhi Al-Harbi (2001), "Application of the AHP in project management", 19(1), pp.19–27.

Muharni, Y., Kulsum, & Khoirunnisa, M. (2019), "Warehouse Layout Designing of Slab Using Dedicated Storage and Particle Swarm Optimization". *IOP Conference Series: Materials Science and Engineering*, 532, 012003.

Ruud H. Teunter; M. Zied Babai; Aris A. Syntetos (2010), ABC Classification: Service Levels and Inventory Costs, 19(3), pp.343–352.

Sawik, T. (2016). A note on the Miller-Tucker-Zemlin model for the asymmetric traveling salesman problem. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 64(3).

Tompkins J. A. (2010). *Facility Planning 4th*, , pp. 574-577

Virendra Rajput, Dinesh Kumar, Arun Sharma, Shubham Singh, Rambhagat, (2018), "A Literature Review on Analytic Hierarchy Process (AHP)"

Zhang, Ren-Qian; Wang, Meng; Pan, Xing (2019), "New model of the storage location assignment problem considering demand correlation pattern". *Computers & Industrial Engineering*, S0360835219300294.

MATHEMATICAL MODELLING IN TRANSPORTATION METHOD APPLICATION FOR RICE OF CAN THO CITY

Nhat Nguyen[1], Anh Duong[2], Ly Tran[3] and Uyen Ngo[4]

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: nhatnguyen.2469@gmail.com[1], dvnanh@hcmiu.edu.vn[2].

Abstract. This paper aims to understand and evaluate the status of the transportation cost of rice from Can Tho City, to distribution points in other provinces and cities as well as major ports in Ho Chi Minh, Ba Ria - Vung Tau. Through learning about transportation costs of countries around the world, then a suitable mathematical model is built to the situation in Vietnam in general and Can Tho city in particular, in order to come up with a reasonable cost. From there, it brings efficiency to the transportation needs in Vietnam. At the same time, there is a benefit to farmers when agricultural products are distributed reasonably at safe prices. This article also discusses and calculates arising problems such as warehouse management, backhaul management, goods preservation and towards the consumption of excess goods, exceeding demand.

Keywords: Rice, Transportation cost, Can Tho City, Mathematical model.

1. Introduction

1.1 Background

1.1.1 Transportation cost

Transportation costs in the logistics industry are all the costs of transporting, circulating and distributing goods from the place of manufacture to the market where the product is sold. This is the biggest cost a customer will pay for any item. Although the authorities and transportation companies have made efforts to cut these costs, prices are still constantly climbing due to fluctuations in fuel prices.

Transportation costs have a great influence on the price of goods in distribution and circulation, the development in production and business of enterprises, the economy of Vietnam and other countries in the world.

In order to compete in the market, Vietnamese products need to promote their own stable prices. However, the current transportation costs in Vietnam are very high compared to the world average, which has a great impact on the prices of domestic products in recent years.

According to the World Bank, transportation costs make up about 59% of Vietnam's logistics costs, and these costs are currently very high, but not just because of the "negative charges" problem. Experts, ministries and branches all acknowledge that transport activities in Vietnam have many shortcomings and the capacity of the entire system is not high.

1.1.2 Can Tho City

Can Tho, the most modern and developed city in the Mekong Delta, is currently a grade I city, an economic, cultural, and social center, as well as a health, education, and commerce hub for the Mekong Delta region.

Can Tho is the capital and nuclear urban area of the Southwestern region since the French colonial period, and continues to be the economic center of the Mekong Delta. In addition to geographical features as an important traffic hub between the provinces in the region, Can Tho city is also known as a river town. The city has a tangled river system, a large area of orchards and fields, famous for Ben Ninh Kieu, Cai Rang Floating Market, a typical cultural activity of the South. According to the plan by 2025, Can Tho city will become the center of industry, trade - services, education - training and science - technology, health and culture of the Mekong Delta. At the same time, it is the gateway urban area of the lower Mekong river basin, is an important hub for intra-regional and international transportation. And will become a fairly developed city in Southeast Asia.

1.2 Problem statements

In a statistic in 2019, 2/3 of the goods in the Mekong Delta still have to be transported to the seaports of Ho Chi Minh City and Cai Mep (Ba Ria - Vung Tau) by road. This results in unnecessary additional costs at very high prices. For example, in road transport costs, gasoline accounts for about 30-35%, average tolls (BOT) accounts for about 10-15%. This is a very large number in the transport process.

Moreover, the organization of transport is not reasonable, while the means of transport in Vietnam are old, with outdated technology, consuming a lot of fuel, lack of large vehicles and specialized vehicles, vehicle cost and investment is high. Handling capacity has a great influence on shipping planning, as limited handling capacity leads to waiting times and increased transportation costs.

Besides, a frequent problem is the overabundance of rice. The production of rice to meet the needs of the market is what everyone wants. However, with natural conditions as well as a large cultivation area, every year, the rice output is always higher than the market demand, leading to an overabundance of rice. Since then, farmers are forced to price, not guaranteed.

Therefore, the application of mathematical models to optimize transportation costs will be applied so that there can be general prices for the cost of transporting rice in the Mekong Delta.

1.3 Objective

Create an algorithmic model to solve the existing problems of the Vietnamese rice market in general, and Can Tho city in particular. Building a transport pricing system as well as creating the best conditions for both businesses and farmers.

With the initial desire to apply to rice in the Can Tho, it is hoped that the application of mathematical models in this field will be more widely applied in Vietnam in many professions as well as in many different industries. Besides, I hope that there will be more optimal and favorable mathematical models for development in Vietnam.

1.4 Supply chain

The purchase of rice from granaries, small traders, and small enterprises is the start of the rice supply chain. There will be two scenarios when getting demand from different destinations. To begin, if the quantity acquired is less than or equal to the quantity demanded, the business owners will analyze and discuss the situation with the requester to see if the full quantity

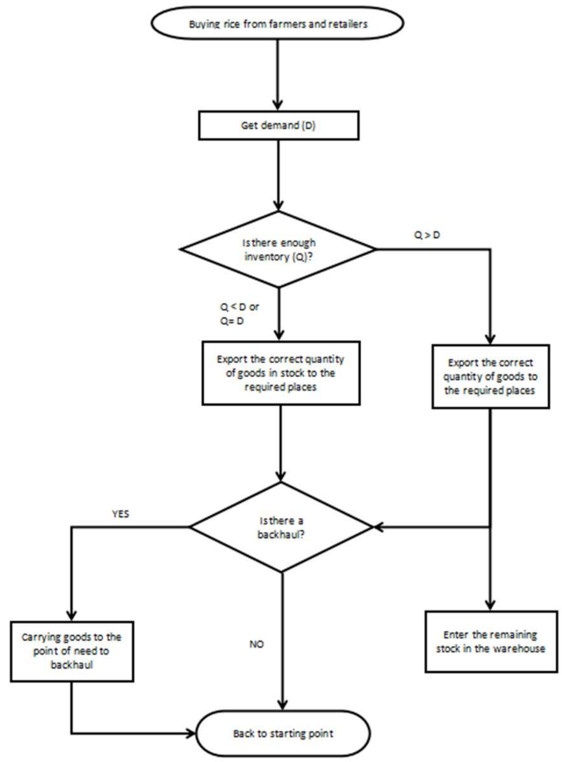


Figure 1: Supply chain of Rice in Can Tho City Design

can be exported. In the second situation, if the quantity in stock exceeds the required quantity, we will send the customer the correct quantity of demand, and the surplus balance will be stored in the warehouse until the next order is received or put into production. Rice products are being developed to cut storage costs and boost sales. Backhaul concerns will be taken into account once the goods have been delivered to their destinations. If this is the case, the products will be delivered to the backhaul point first, followed by a return to the beginning point.

2. Literature Review

The high cost of transportation also partly affects the overall cost of Logistics. According to Current Status and Solutions to Reduce Logistics Costs In Vietnam (Nguyen Hoang Phuong, 2019).

> Logistics costs = Transportation costs + Capital opportunity cost for inventory + Cost of preserving goods.

According to the formula, higher transportation costs lead to higher logistics costs. Therefore, the stabilization of these costs is essential. Logistics cost is an extremely important fee to consider. Many businesses in Vietnam currently do not take advantage of logistics; in fact, some businesses do not recognize the critical role of logistics in lowering production costs. In Vietnam, commodities must pass through so many intermediaries, from the procurement of raw materials to manufacturing and delivery to customers, resulting in higher transaction costs and selling prices. Because of the lack of knowledge, the members of the chain only know that they have a direct connection with their companies and are unaware of the other members, logistics costs rise. The distribution system, on the other hand, is mostly focused on urban areas, but rural areas are mostly ignored.

It is essential to create mathematical models to simplify problems. To be able to form, we still have to grasp the basic information of the object, the target, then we can build accordingly. Not just about any mathematical model can be applied anywhere. The paper provides the first quantitative model and optimistic analysis on the effect of the Thanh Hai-Tay railway, similar to the article Effect of Transportation Cost on Logistic Transaction Efficiency and Outsourcing - with Qinghai-Tibet Railway as an example, Hongchang Li and Chaohe Rong (2006), which was very good in identifying transportation costs as the heart. The system is set up to ensure that transaction logistics are as effective as possible. The results show that since the system went live in July 2006, the quality of logistics transactions has significantly improved. This model, however, cannot be fully implemented in the Delta because it is designed for rail transportation, which is not widely used in the Can Tho.

The importance of transport in agriculture is irrefutable. The improvement of the transport system will greatly contribute to the productivity and quality of agricultural products. The impact of transportation on agricultural production in a developing country: a case of kolanut production in Nigeria, Ajiboye, A. O. and O. Afolayan (2009) considers transportation to be a critical factor in increasing agricultural production. It people's quality of life, opens new main economic regions, and creates markets for agricultural products. It also encourages parent between geographical and economic regions. Improved transportation, according to this report, would enable farmers to work harder in rural areas in order to increase production, add value to their goods, reduce spoilage and wastage, motivate farmers, and have a positive effect on their productivity, income, jobs, and poverty levels in rural areas. With the survey and the success of the article, we can see the huge impact of transport on the farmer as well as on the market.

Not only does transportation cost economic, but transportation cost also affects volunteering. According to article Mathematical Modelling On Transportation Method Apllication For Rice Distribution Cost Optimization, Azizah (2019), the organization had to sort and calculate the operations to be able to optimize the costs incurred by transportation cost. Since then, a reasonable mathematical model has been built and has successfully solved this difficult problem.

$$Minimumkan\ Z = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij}$$

Constrain function

$$\sum_{i=1}^{m} C_{ij} X_{ij} = a_i \ ; i = 1, 2, \ldots \ldots, m$$

$$\sum_{j=1}^{n} C_{ij} X_{ij} = b_j \ ; j = 1, 2, \ldots \ldots, n$$

Figure 2: Mathematical Model of Azizah (2019)

In general, the above studies and algorithms have partly improved the situation of that locality or organization. The application and improvement of the mathematical model from the above references to match will be considered more carefully in this article.

2. The Optimization Model

3.1 Parameters

After careful consideration and selection, Linear programming is the most effective and optimal method for the problems encountered here.

C_{ij}: Transportation cost per ton per kilometer of goods from source i to destination

- a_i: Quantity of goods offered or capacity from source i
- b_j: Quantity of goods requested or ordered by destination j
- D_{ij}: Demand per ton of the destination j to source i
- K_{ij}: Percentage of damage lost of good in the trip from source to destination j
- B_{ji}: Backhaul profit when truck comes from destination j back to source i
- F_{ij}: Fixed cost inventory needed when the trip from source i to destination j
- H_{ij}: Holding cost inventory needed when the trip from source i to destination j

- P_{ij}: The product in tons that source i bought before transport to destination j

3.2 Decision Variable

- O_{ij}: Exported price from source I to destination j
- X_{ij}: Number of goods in ton distributed from source i to destination j
- Y_{ji}: Number of goods in ton distributed backhaul from destination j back to source i
- $V1_{ji}$: equals 1 if backhaul appears in the trip from destination j back to source i. Equals 0 if does not have backhaul
- $V2_{ij}$: equals 1 if the product in tons that source i bought before transport to destination j is greater than number of goods in ton distributed from source i to destination j. Equals 0 if product in tons that source i bought before transport to destination j is smaller than number of goods in ton distributed from source i to destination j.

3.3 Mathematical model

With the application of the knowledge from the main refenrence, the model of the main reference is as follows:

$$\textbf{Maximize Z} = \sum_{i=1}^{m}\sum_{j=1}^{n} O_{ij}X_{ij} - \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}X_{ij} - \sum_{i=1}^{m}\sum_{j=1}^{n} K_{ij}X_{ij} + \sum_{j=1}^{n}\sum_{i=1}^{m} 0.03 * B_{ji}Y_{ji} - \sum_{i=1}^{m}\sum_{j=1}^{n}(V_{ij} * (F_{ij} + ((P_{ij} - X_{ij}) * H_{ij})))$$

To easy, the mathematical model is divided in to five part:

- $A * = \sum_{i=1}^{m}\sum_{j=1}^{n} O_{ij}X_{ij}$: the revenue of export rice from source i to destination j.
  $B * = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}X_{ij}$: the normal transportation cost from the source i to destination j.
- $C \ * = \sum_{i=1}^{m}\sum_{j=1}^{n} K_{ij}X_{ij}$: the damage cost when transported (damp, moldy,…).
- $D \ * = \sum_{j=1}^{n}\sum_{j=1}^{m} 0.03B_{ji}Y_{ji}$: the backhaul profit is able to get.
- $E \ * = \sum_{i=1}^{m}\sum_{j=1}^{n}(V1_{ij} * (F_{ij} + ((P_{i} - X_{ij}) * H_{ij})))$: the inventory cost when overstock.

The model can be rewritten in this way:

$$\text{Max Z} = A * - B * - C * + D * - E *$$

3.4 Constraints

In order to best optimize the mathematical model, setting constraints is a must. These conditions will follow throughout the mathematical model. From there, the results can give us the most optimal choices.

There are 8 constraints in this mathematical model.

$$\sum_{i=1}^{m} \boldsymbol{X_{ij}} <= \boldsymbol{a_i},\ \forall i = 1,2,\dots m \quad (1)$$

The quantity of goods transported must be less than or equal to quantity of goods offered or capacity from source i.

$$\sum_{j=1}^{n} \boldsymbol{X_{ij}} <= \boldsymbol{b_j},\forall j = 1,2,\dots n \quad (2)$$

The quantity of goods transported must be less than or equal to quantity of goods offered or capacity from destination j.

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{K_{ij}X_{ij}} < \sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{C_{ij}}\,\boldsymbol{X_{ij}},\ \forall i = 1,2,\dots m \text{ and } j = 1,2,\dots n \quad (3)$$

The cost incurred due to damage to the goods must be less than the value of the goods multiplied by the amount of freight.

$$\mathbf{Y_{ji}} <= \boldsymbol{X_{ij}},\ \forall i = 1,2,\dots m \text{ and } j = 1,2,\dots n \quad (4)$$

The quantity of goods transported backhaul must be less than or equal to the quantity transported to the destination.

$$\boldsymbol{V_{ij}} <= \mathbf{1},\forall i = 1,2,\dots m \text{ and } j = 1,2,\dots n \quad (5)$$

This is a binary value, showing only one of the values 0 and 1. If 1 occurs this shipment has inventory for stock, and vice versa, if 0 occurs, this shipment has no inventory for stock.

$$\boldsymbol{X_{ij}} <= \boldsymbol{D_{ij}},\forall i = 1,2,\dots m \text{ and } j = 1,2,\dots n \quad (6)$$

The shipping volume must always be less than or equal to the order's deal.

$$P_i - X_{ij} < V1_{ij} * bigM, \forall i = 1,2,\dots m \text{ and } j = 1,2,\dots n \quad (7)$$

$$X_{ij} - P_{ij} >= V1_{ij} * bigM, \forall i = 1,2,\dots m \text{ and } j = 1,2,\dots n \quad (8)$$

The two conditions number 7 and number 8 above are intended to limit the binary factor. At the same time, it also tells us whether the number of products in stock is greater or less than the quantity that needs to be shipped.

3.5 Solution development

To solve this mathematical model, ILOG CPLEX is used to find out the solution. ILOG CPLEX is optimization software, whose name derives from the simplex method and the C programming language, although today the software uses the point-in method and uses both languages. C++, C# and Java. This software can solve integer programming problems, extremely large linear programming problems, quadratic programming problems and convex programming problems.

With the above mathematical model described, simulated a shipment of enterprises in Can Tho to 2 important export destinations: Cat Lai Port (Ho Chi Minh City) and Cai Mep Port (Ba Ria - Vung Tau). These are the two largest deep-water ports in the South of Vietnam, and also an important waterway traffic hub of the country as well as the world.

By figuring out how much should be shipped, how much should be backhaul, and how much inventory should be in stock, the actual total cost in addition to normal shipping costs can be calculated. Through that, it is possible to know more accurately the costs incurred, the added values, thereby improving the transport system of enterprises to bring the highest economic efficiency.

3.6 Code

** Model*

```
int numsr=...;
int numgl=...;
int bigM=10000000000;

range source=1..numsr;
range goal=1..numgl;

float O[source][goal]=...;
float C[source][goal]=...;
float a[source]=...;
float b[goal]=...;
float D[source][goal]=...;
float K[source][goal]=...;
float B[source][goal]=...;
float F[source][goal]=...;
float H[source][goal]=...;
float P[source]=...;


dvar int+ X[source][goal];
dvar int+ Y[source][goal];
dvar boolean V[source][goal];

maximize
  sum(m in source, n in
goal)(O[m][n]*X[m][n])

- sum(m in source, n in
goal)(C[m][n]*X[m][n])

- sum(m in source, n in
goal)(K[m][n]*X[m][n])

+ sum(m in source, n in
goal)(0.3*B[m][n]*Y[m][n])

- sum(m in source, n in
goal)(V[m][n]*(F[m][n]+(P[m]-
X[m][n])*H[m][n]));

subject to {

forall (m in source, n in goal){
    sum(m in source)X[m][n]<=a[m];
}

forall (m in source, n in goal){
    sum(n in goal)X[m][n]<=b[n];
}

forall (m in source, n in goal){
    sum(m in source, n in
goal)K[m][n]*X[m][n]<= sum(m in
source, n in goal)C[m][n]*X[m][n];
}


forall (m in source, n in goal){
    V[m][n] <= 1;
}

forall (m in source, n in goal){
    X[m][n] <= D[m][n];
}



forall (m in source, n in goal){
```

```
    P[m] - X[m][n] <= V[m][n]*bigM;
    X[m][n] - P[m]>= V[m][n]*bigM;
}

forall (m in source, n in goal){
    X[m][n]>=Y[m][n];
}

forall (m in source){
    P[m] <= a[m];
}
}
```

** Data*

```
numsr=41;
numgl=2;

SheetConnection DATA("DATA1.xlsx");
O from SheetRead(DATA,"Pro");
C from SheetRead(DATA,"CC");
a from SheetRead(DATA,"a");
b from SheetRead(DATA,"b");
D from SheetRead(DATA,"Demand");
K from SheetRead(DATA,"Dam");
B from SheetRead(DATA,"BH");
F from SheetRead(DATA,"FC");
H from SheetRead(DATA,"HC");
P from SheetRead(DATA,"PUR");

SheetConnection
RESULT("DATA1.xlsx");
X to SheetWrite(RESULT,"Xij");
Y to SheetWrite(RESULT,"Yij");
```

4. Result analysis

The output results show us the quantity of goods that should be exported to the ports. Through the results, the priority of cargo to Cat Lai port is higher than that of Cai Mep. Specifically, in the position of the 22nd enterprise, they prioritize transporting their goods to Cat Lai more than to Cai Mep. The reason is due to the long distance traveled, the cost of petrol and BOT, but the return value is not much higher. Not to mention the labor and time are also greatly affected.

In addition, if the backhaul is given the highest priority, the number of backhauls is equal to the number of goods shipped, it will optimize profits and minimize development costs as much as possible. Because if goods are only carried on one way, it is extremely wasteful for the empty truck to return to the starting point.

Table 1. Preliminary results

Enterprise	X_{i1} (tons)	X_{i2} (tons)	Y_{i1} (tons)	Y_{i2} (tons)
1	34	0	34	0
2	32	0	32	0
3	36	0	36	0
4	0	38	0	38
5	22	0	22	0
6	27	0	27	0
7	37	0	37	0
8	27	0	27	0
9	42	0	42	0
10	26	0	26	0
11	51	0	51	0
12	0	28	0	28
13	39	0	39	0
14	27	0	27	0
15	20	0	20	0
16	36	0	36	0
17	0	37	0	37
18	26	0	26	0
19	20	0	20	0
20	42	0	42	0
21	40	0	40	0
22	31	15	31	15
23	37	0	37	0
24	39	0	39	0
25	26	0	26	0
26	0	47	0	47
27	37	0	37	0
28	34	0	34	0

29	42	0	42	0
30	33	0	33	0
31	20	0	20	0
32	42	0	42	0
33	38	0	38	0
34	25	0	25	0
35	0	36	0	36
36	35	0	35	0
37	25	0	25	0
38	25	25	25	25
39	0	35	0	35
40	44	0	44	0
41	27	0	27	0

4.1 Total revenue for the transport (A*)

Ignore calculating (total revenue for the transport process) because A* is only conditional for this mathematical model. Ignore calculating A* (total revenue for the transport process) because A is only conditional for this mathematical model. Through A*, the mathematical model is easier to implement in the form of optimal profit, minimizing costs.

4.2 The normal transportation cost (B*)

Based on the output from the mathematical model, we can calculate the normal transportation cost. Usually, this is the factor that Vietnamese businesses are currently interested in. They only notice how much one of their shipments costs in total. Therefore, if businesses look at this result, they will think that it is quite high compared to the common ground of about 18 to 25 million for a shipment.

4.3 Damage cost (C*)

The cost of damage is one of the small costs in the transportation of goods, so businesses often ignore this cost. Because it only ranges from 1.5% to 3.5%, this cost is not significant. However, if the quantity of goods is not carefully managed, this high cost is entirely possible.

4.4 Backhaul profit (D*)

This is one of the profits that businesses forget. Usually because this only occurs at 30% probability, businesses do not care about this amount of money. However, if viewed objectively, this is not a small amount. If businesses arrange and increase the number of backhaul deliveries even more, it will greatly improve transportation costs.

4.5 Inventory cost (E*)

This is an unexpected expense for businesses. Because they can't predict demand, and when the season comes, they always buy first from farmers, so when they receive demand, the excess stock will have to be carefully preserved for later deliveries. If overall a shipment like this, this cost is really not high, however, if it is a whole season, the cost for this storage will increase many times.

4.6 Total transportation cost for a trip

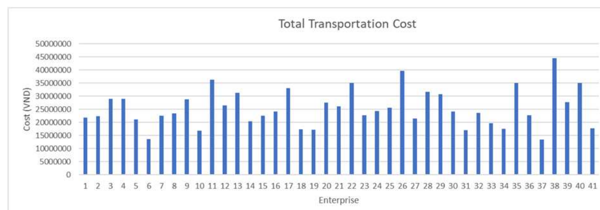


Figure 3: Total transportation cost

In fact, if the overall impact factors are taken into account, the cost of transport at present is quite high compared to the common ground. The cost can range from 15 to more than 40 million depending on the quantity of goods transported. However, when applying and properly calculating the costs, businesses can reduce the number of unnecessary costs such as storage costs, damage costs, etc. reduce the load, the transportation cost will be reduced.

1. Conclusion

This thesis has introduced factors affecting the cost of rice transportation such as spoilage costs, storage costs from Can Tho to major ports in the country. From there, a mathematical model is formed to be able to calculate the most optimal quantity of transported goods to minimize transportation costs.

With this mathematical model, if applied optimally, businesses will be able to calculate the real costs of the transportation process, no longer depending on intermediaries. From there, autonomously give their businesses the transportation problems.

The application of mathematical models to reduce transportation costs not only supports businesses but also helps farmers stabilize prices.

However, the mathematical model still has many limitations when it is applied and applied to only one transportation. Hopefully in the future, this model will be upgraded or replaced to be more suitable for practice. Moreover, the further application of technology to directly calculate the incurred costs will be applied, instead of only temporary calculation as in the current mathematical model.

Finally, calculations aimed at increasing costs such as backhaul should be considered in the future. Furthermore, the handling of inventory is also an issue that businesses should look forward to in the future.

Reference

Ajiboye, & Afolayan, O.T. (2010). The impact of transportation on agricultural production in a developing country : a case of kolanut production in Nigeria.

Azizah, N. (2019). Mathematical Modelling On Transportation Method Apllication For Rice Distribution Cost Optimization. *CAUCHY*, *5*(4), 195. doi: 10.18860/ca.v5i4.4893

Francesc B. M., (2019). "An optimization model for multi-deep storage ". Department of Industrial Systems Engineering and Product Design, the Polytechnic University of Catalonia, UPC.

Hongchang Li, & Chaohe Rong. (2008). Effect of transportation cost on logistic transaction efficiency and outsourcing - with Qinghai-Tibet railway as an example. *2008 IEEE International Conference On Service Operations And Logistics, And Informatics*. doi: 10.1109/soli.2008.4683033

Muppani, V. R., & Adil, G. K. (2008*)*. "A branch and bound algorithm for class-based storage location assignment". *European Journal of Operational Research,* 189(2), 492–507.

Muther R. and Hales L. (2015). "Part One, Part Two and Part Three*", Systematic Facility Layout Planning"*, 4th .ed, pp ix-13-1.

Phuong, N. (2019). CURRENT STATUS AND SOLUTIONS TO REDUCE LOGISTICS COSTS IN VIETNAM. *Malaysian E Commerce Journal*, *3*(3), 01-04. doi: 10.26480/mecj.03.2019.01.04

Rodoljub V., Nenad M., Snezana V. (2016). Dynamic Storage Systems. *ANNALS of Faculty Engineering Hunedoara – International Journal of Engineering*.

SPLIT DELIVERY VEHICLE ROUTING PROBLEM WITH TIME WINDOWS

Nguyen Le Thanh Thuy[1], *Duong Vo Nhi Anh[2], Do Vinh Truc[3], Tran Van Ly[4]

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: [1]nglethanhthuy227@gmail.com, [2]dvnanh@hcmiu.edu.vn, [3]dvtruc@hcmiu.edu.vn,[4]tvly@hcmiu.edu.vn

Abstract: In the field of distribution, routing optimization for a fleet of vehicles to service a set of customers is an important factor to reduce transportation expenses for logistics or transport companies. However, any logistics company has its own problems that need to be addressed. Many 3PL company is facing difficulties in delivering goods on time as specified by customers. The number of customers is large, but the route is not predetermined, it is time- consuming because the distance traveled is too long. Moreover, when the customer's demand exceeds vehicle's capacity, it is necessary to visit a customer more than one trip. The purpose of this thesis is to find optimal route by minimizing the travelling time without violating some constraints such as the capacity of the truck, the time window of the customer. An integer linear programming of the Split Delivery Vehicle Routing Problem with Time windows (SDVRPTW) is presented in this paper. Although the overall computational time was relatively lengthy, the results indicated that the produced solution was better than the existing routing and scheduling that the company used.

Keywords: Split Delivery Vehicle Routing Problem with Time Windows (SDVRPTW), Vehicle Routing Problem (VRP), Split load delivery, Delivery with Time Window, Third-party logistics.

1. Introduction

Supply Chain Management can be defined as "the management of upstream and downstream relationships with suppliers and customers in order to create enhanced value in the final marketplace at less cost to the supply chain as a whole". Supply chain management involves collaboration between firms to connect suppliers, customers, and other partners as a means of boosting efficiency and producing value for the end consumer (as explained by Michigan State University professors in the text Supply Chain Logistics Management). Having the right product at the right time at the right place and in the right condition – these are the well-known requirements for logistics and transportation in general. But fulfilling these requirements is getting more and more complex in a dynamically changing logistic environment. In order to be competitive on the market, any company must improve its supply chain management in order to satisfy its customers at the best service level and the lowest possible cost. It is therefore of key importance to optimize the logistics flow at different levels: strategic, tactical and operational. Moreover, profits are one of the key factors that evaluate the success of Logistics enterprises, in which logistics costs are also the main key in their competitive strategy. Therefore, the cost control is an effective measure to help 3PLs compete better. Successful businesses need an accurate picture of demand to drive production, inventory, distribution, and buying plans across their operations. These challenges are intensified by the effects of seasonality, promotions, and product proliferation, not to mention growth through mergers and acquisitions.

For large companies, owning a large number of customers is obvious. However, it is hard to answer the question "What is the most optimal way to transport goods to the customer's location?". With the large of orders that need to be serviced daily, each vehicle must handle more than two customers per their route at full load to optimize vehicle load. Moreover, there are orders which have demands exceed vehicle's capacity. Therefore, it is necessary to visit customer's location by multiple trips. However, the company does not have a plan to arrange and offer the most optimal routes, which are divided based on the criteria of prioritizing groups of near customers for drivers. Another factor that needs to be considered is the time window of each customer. Accordingly, vehicles arriving at customer warehouses outside of this period will not be allowed to deliver. Traveling distance is too long, causing time- consuming and affecting delivery on time. Unsuccessful deliveries will reduce KPIs as well as transportation profits for the company.

This study is inspired by the real case of vehicle routing optimization in a 3PL Company. The aim of this research is to find the optimal routing for delivering products by minimizing the total travelling distance and cost assumption and increasing customer satisfaction by optimize the delivery time and meet customer's demand on time by solving the problem with a set of locations, in a route contains all locations, on conditions, that route must begin and end at the depot, all customers must be fulfilled the demand by single/ multiple trips based on the volume of order and satisfied time window of customers. In this thesis, there are a few of customers who has the volume of order is larger than vehicle's capacity. Therefore, it is necessary to delivery by multiple trips to meet customer's demand. The order is completed when the total delivered load in trips for that customer is equal to the demand that customer ordered. Moreover, an important factor needs to be considered is time windows. Distribution process is considered complete when the demand is delivered completely and does not violate the time windows specified by the customer. The split delivery vehicle routing problem with time windows (SDVRPTW) is employed to solve the problem.

In mathematics, computer science and economics, an optimization problem is the problem of finding the best solution from all feasible solutions. In the Vehicle Routing Problem (VRP). Split Delivery Vehicle Routing Problem (SDVRP) is one of variants of VRP, meaning that the customer's demand can be split among several vehicles on different routes. The SDVRP was introduced by Dror (1989) and Trudeqau (1994) who defined the problem, derived some structural properties, and proposed a local search heuristic. They empirically demonstrated in their paper that allowing split deliveries can result in savings that are significant both in the total distance traveled and the number of vehicles required. However, with the split deliveries, the issue is exacerbated. According to Damon Gulczynski (2010), it is especially undesirable for the customer to be interrupted and distracted twice (or more), unless the delivery is substantial in amount or value each time. Therefore, in their paper, they allowed a customer's demand to be split among several vehicles only if each vehicle delivers at least a minimum amount when it visits the customer. They used a modified Clarke-and-Wright savings algorithm to address this problem. WS Liu (2012) and T Xiang (2016) also proposed a clustering algorithm which arranges paths after grouping was proposed to solve this problem. According to this paper, the customers who load were greater or equal to the vehicle load limit was arranged in advance. Then combined with the distance between customers and load, a split threshold was set to limit the load of vehicle to a certain range. According to the nearest principle, all customers were clustered and grouped. If the customer load in a group does not reach the minimum load of vehicle and is beyond the limit load when new customers are added in, the new customers were split and adjusted. Finally, while all the customers were divided into groups, the customers paths for each group were arranged by Ant ColonyOptimization (ACO) algorithm. Accordingly, the customer's demand exceeds the vehicle's capacity, it is necessary to have more than one visit per customer in order to fully fulfill the order. Furthermore, each customer has a fixed time window, mean that the delivery to a customer can be split between any number of vehicles and must to meet the time window of customers. According to Pierre Dejax and Michel Gendreau (2002), the Split Delivery Vehicle Routing Problem with Time Windows (SDVRPTW) consists in determining a least cost set of vehicle routes such that every route starts and ends at the depot and that every customer is served by one or several vehicles, the vehicle capacity being respected and the service of customers beginning inside their time windows. In their paper, they used a Branch and Price approach for solving the SDVRPTW without imposing restrictions on the split delivery options. Besides that, in the paper of Nicola Bianchessi and Michael Drexl (2019), on the customer side, though, several visits cause inconvenience, as at each visit, the customer has to interrupt his primary activities to handle the goods receipt. They introduced generalizations of the SDVRP that allow to control the degree of inconvenience caused by split deliveries and to balance overall distribution costs and customer satisfaction. This creates a win-win situation for transport companies and their customers by designing an extended branch and cut algorithm and defining the corresponding generalization that considers customer inconvenience constraints to solve the SDVRPTW-IC. E Latiffianti (2017) applied this problem in a case study involving pickups and deliveries of workers from several points of origin and several destinations by using an exact method. Each origin represents a bus stop, and the destination represents either site or office location. The produced solution was also capable of reducing fuel cost by 9% that was obtained from shorter total distance travelled by the shuttle buses. Splitting also allows to decrease costs according to S.C. Ho, D. Haugland (2004). They also proposed a solution method based on tabu search for solving the VRPTWSD without imposing any restrictions on the split delivery options.

In this paper, K-Nearest Neighbor Algorithm (KNN) is proposed to solve Split Delivery Vehicle Routing Problem with Time Windows (SDVRPTW). The nearest neighbor algorithm (NN) was one of the first algorithms used to solve the travelling salesman problem approximately. In that problem, the salesman starts at a random city and repeatedly visits the nearest city until all have been visited. The algorithm quickly yields a short tour. One of the simplest decision procedures that can be used for classification is the nearest neighbor rule. It classifies a sample based on the category of its nearest neighbor. When large samples are involved, it can be shown that this rule has a probability of error which is less than twice the optimum error - hence there is less than twice the probability of error compared to any other decision rule. he simple version of the K-nearest neighbor classifier algorithms is to predict the target label by finding the nearest neighbor class.

The rest of the paper is organized as follows. The next section presents the mathematical model for SDVRPTW while the third section describes the framework for apply KNN algorithm to solve the large-scale problem. Finally, the computational result and conclusion are presented in the final section.

2. Mathematical Model

The model is constructed by giving a directed graph including customer set as vertices set

Customers: The problem is given a set of customers C = {0, 1, …, n}, residing at n different locations. Denoting that node 0 representing the location of the depot, N = C ∪ {0} become the set of all nodes are considered in this problem. Every pair of locations (i, j), where i, j ∈ N and i ≠ j, is associated with a cost of travelling d_{ij} and a traveling time t_{ij}. Each customer i ∈ C has a demand $D_i > 0$. For each trip, it has a travelling cost c_T per distance unit.

Trips: A set of trips T = {1, …, max $(\frac{Demand\ of\ each\ customer\ i}{Capacity\ of\ vehic\ \ k})$} which is delivered by a fleet of vehicles with identical capacities Ca, each trip has a fixed cost c_F for a vehicle.

Time windows: Each customer i ∈ C has a time window, i.e. an interval $[e_i, l_i] \subseteq \Re$, where e_i and l_i are the earliest and latest time to start to service customer i. A vehicle may arrive at customer location i before e_i, but cannot start servicing until the time window opens at e_i. A vehicle cannot arrive at customer location i after the time window closes at l_i.

Split deliveries: The demand of a customer may be fulfilled by more than one vehicle. This occurs in all instances where some demand exceeds the vehicle capacity but can also turn out to be cost effective in other cases.

For each arc (i, j), where i, j ∈ N and i ≠ j and for each trip t, we defined x_{ij}^t as $x_{ij}^t = 1$ if trip t is delivered directly from customer i to customer j and 0 otherwise.

The decision variable u_i^t is the total quantity of goods delivery to customer i by trip t, decision variable r_{it} denotes the time the trip t starts to service customer i and q_{ij}^t is the quantity of goods when the vehicle travels directly from i to j in trip t. The last decision variable is f_i which means final service time of customer. This variable must set to determine the last time start to serve customer i in order to ensure order fulfillment without violating the customer's time window.

The split delivery vehicle routing problem with time windows from depot to customers can be formulated as follows:

$$Z = \min \left[c_T * \left(\sum_{t\in T}\sum_{i\in N}\sum_{j\in C} d_{ij}x_{ij}^t + \sum_{t\in T}\sum_{i\in C}\sum_{j\in N} d_{ij}x_{ij}^t \right) + c_F * \sum_{t\in T}\sum_{i\in C} x_{0i}^t \right] \quad (1)$$

Subject to

$$\sum_{\substack{i\in N\\ i\neq j}} x_{ij}^t \leq 1 \qquad \forall j \in C, \forall t \in T \qquad (2)$$

$$\sum_{\substack{j\in N\\ j\neq i}} x_{ij}^t \leq 1 \qquad \forall i \in C, \forall t \in T \qquad (3)$$

$$\sum_{\substack{i\in N\\ i\neq h}} x_{ih}^t - \sum_{\substack{j\in N\\ h\neq j}} x_{hj}^t = 0 \qquad \forall h \in C, \forall t \in T \qquad (4)$$

$$\sum_{j\in C} x_{0j}^t = \sum_{j\in C} x_{j0}^t \qquad \forall t \in T \qquad (5)$$

$$\sum_{\substack{i\in N\\ i\neq h}} q_{ih}^t = u_h^t + \sum_{\substack{j\in N\\ j\neq h}} q_{hj}^t \qquad \forall h \in C, \forall t \in T \qquad (6)$$

$$\sum_{i\in C} q_{0i}^t = \sum_{i\in C} u_i^t \qquad \forall t \in T \qquad (7)$$

$$\sum_{t\in T} u_i^t = D_i \qquad \forall i \in C \qquad (8)$$

$$q_{ij}^t \leq Ca * x_{ij}^t \qquad \forall i \in N, \forall j \in C, \forall t \in T \qquad (9)$$

$$q_{i0}^t = 0 \qquad \forall i \in C, \forall t \in T \qquad (10)$$

$$x_{ji}^t \leq u_i^t \qquad \forall i \in C, \forall j \in N, \forall t \in T \qquad (11)$$

$$\sum_{i\in C} x_{0i}^t \leq K \qquad \forall t \in T \qquad (12)$$

$$r_{it} + t_{ij} + s_i - BigM*(1 - x_{ij}^t) \leq r_{jt} \qquad \forall i \in N, \forall j \in N, \forall t \in T \qquad (13)$$

$$f_i = \max(r_{it}) \qquad \forall i \in N, \forall t \in T \qquad (14)$$

$$e_i \leq f_i \leq l_i \qquad \forall i \in N \qquad (15)$$

$$x_{ii}^t = 0 \qquad \forall i \in N, \forall t \in T \qquad (16)$$

$$r_{it} \geq 0 \qquad \forall i \in C, \forall t \in T \qquad (17)$$

$$x_{ij}^t \in \{0, 1\} \qquad \forall i, j \in N, \forall t \in T \qquad (18)$$

$$e_i \geq 0 \ \&\ l_i \geq 0 \qquad \forall i \in N \qquad (19)$$

Constraint (1): The objective of function is to minimize total transportation cost including the cost of moving from the depot to the customer and vice versa and the fixed cost of renting a vehicle.

Constraint (2), (3) and (4): State that each vehicle is operated in a continuous route, which starts

from a depot, visits customers, then returns to the depot.

Constraint (5): Ensure that the delivery must be served by the same trip.

Constraint (6): Relate the quantity of goods at nodes with their succeeding ones on the route.

Constraint (7): Ensure that the quantity of delivered goods must be equal to the total quantity at customer i by trip t.

Constraint (8): Make sure that the customer's demand will be fulfilled.

Constraint (9): Guarantee that the quantity delivered by each trip does not exceed the vehicle capacity.

Constraint (10): Returning the depot with no quantity, means that the goods must be delivered in full.

Constraint (11): Ensure that the trip will not be made to customer who has not demand.

Constraint (12): Guarantee the vehicle resources.

Constraint (13): A vehicle cannot arrive at customer j before $r_{it} + t_{ij} + s_i$ if it is traveling directly from customer i to customer j in trip t.

Constraint (14): To determine the last service time at customer i.

Constraint (15): Time windows constraint.

Constraint (16): Make sure trip t has to leave after serving customer i.

Constraint (17), (18) and (19): The range of decision variables.

3. Solution Approach

The K-Nearest Neighbor algorithm is used to find a near-optimal solution. In order to minimize the total travel cost, we construct a solution by making effort to fully utilize a vehicle over the day (minimize the total number of vehicles used). It means that when a vehicle cannot be used any more, a new route with a new vehicle is created. The main procedures of the proposal algorithm for SDVRPTW are illustrated in the flowchart of Figure 1.

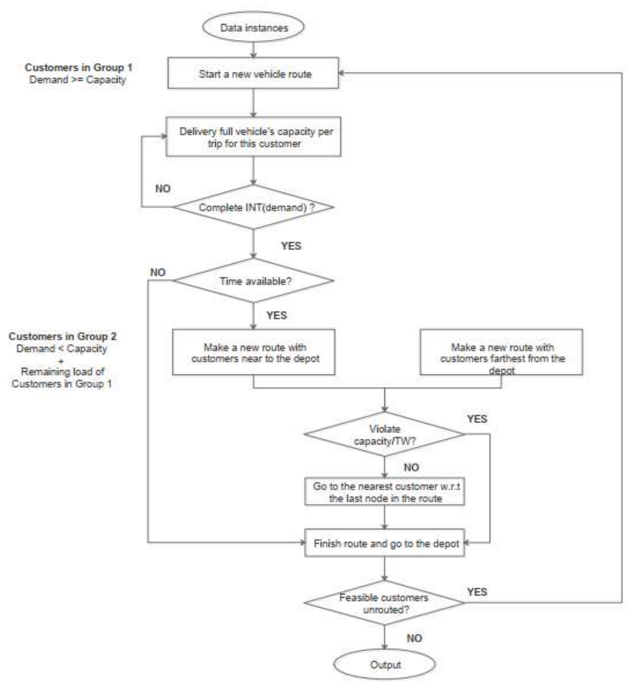


Figure 1. The detail flow chart K-Nearest Neighbor Algorithm for SDVRPTW

Generate solution

First, we will proceed to divide the list of customers into two groups. The first group is the group of customers whose demand exceeds the vehicle's capacity. The second group is all remaining customers, which means that the group of customers whose demand does not exceed the vehicle's capacity.

For Customers in Group 1 (Customer's demand > Vehicle's capacity): Deliver the integer number of customer's demand

We choose a customer who has largest demand and assign for truck 1. Each shipment will be delivered by full capacity of vehicle an integer number of vehicle's capacity. Due to customer's demand exceed vehicle's capacity, it is necessary to visit each customer by multiple trips. Note that the number of trips of each customer will be determined by dividing customer's demand for vehicle's capacity and then rounding up. Moreover, the travelling distance from the depot to customer is the same for each trip and must be done by multiple trips, it is more efficient to deliver by one vehicle instead of multiple vehicles which will incur a fixed cost for renting a vehicle. We will arrange vehicle to deliver goods to this customer group with full capacity for each shipment, and all trips will be delivered by one vehicle until this vehicle cannot be used any more, meaning that this vehicle cannot meet the customer's time windows. Then, a new vehicle is created and start to be used at the same time with the first vehicle. Delivery to this customer group is considered successful until the integer number of customer's demand is delivered, i.e, demand of customer 1 equal to 416 cartons and the vehicle's capacity is 200 cartons. Therefore, it needs 3 trips to fulfill customer's demand in full. As mentioned above, one vehicle will be assigned to deliver 2 trips with full capacity to this customer, meaning that 400 cartons of customer 1 have been completed by 2 trips. The remaining demand of this customer will be solved in the next group. It will reduce the number of vehicles use as well as minimize the cost for renting vehicles.

For Customers in Group 2 (Customer's demand <= Vehicle's capacity + Remaining demand of Customers in Group 1)

In this group, the demand of customer is smaller than vehicle's capacity. It can combine the order of multiple customers for delivering to fulfill all customer's demand. Because the customer has a time window, we will choose a customer in this group that is located farthest from the depot to meet the customer's needs without violating the time window. Then, nearby customers will also in turn be combined into the same route to optimize the delivery process. Customers will be added to the route one by one until the total demand of all customers in the same route does not exceed the vehicle's capacity. Similar to Group 1, a vehicle will be fully utilized until it cannot be used anymore, and a new vehicle will be assigned for other routes at the same starting time with the first vehicle.

In addition, we can make optimal use of the number of vehicles which still have time left after completing the customer group in group 1. However, this is possible if and only if the remaining time is of those vehicles must ensure the time window of the customer in group 2. Therefore, these cars will be assigned to the customers in group 2 who are closest to the depot.

Repeating until the demand of all customers in both groups are satisfied and does not violate the customer's time window.

Following is the description of K-Nearest Neighbor Algorithm for Split Delivery Vehicle Routing Problem with Time Windows (KNN for SDVRPTW) (Algorithm 1).

Algorithm 1: KNN algorithm for SDVRPTW

```
Generate the Initial Solution
    sort number of customers in decreasing of demand
    foreach iteration do
    repeat
        foreach vehicle do
            repeat
                choose customer has the largest demand
                    repeat
                        delivery with full vehicle's capacity
                        add the new routes into set of solution
                    until complete this customer's demand
                    delete this customer from unrouted customer set
            until complete group of customers have demand exceed vehicle's capacity
            choose customer has the nearest distance from depot
            repeat delivery the whole demand of this customer
                select the next customer by apply Nearest Neighbor Algorithm
                if the next node satisfies all constraints then
                    add the next customer into the route
            until the next customer violates vehicle's capacity constraint
            add the new routes into set of solution
            delete these customers from unrouted customer set
            choose customer has the furthest distance from depot
            repeat delivery the whole demand of this customer
                select the next customer by apply Nearest Neighbor Algorithm
                if the next node satisfies all constraints then
                    add the next customer into the route
            until the next customer violates vehicle's capacity constraint
                add the new routes into set of solution
                delete these customers from unrouted customer set
        if this vehicle violates time window constraint then
            Make a new route with new vehicle
            Add all solutions to the solution set
    until all customers are routed
Feasible checking
Output: The set of the best solution
```

4. Computational Results and Analysis

This section is devoted to the experimental evaluation of K-Nearest Neighbor Algorithm on the real case study. In section 4.1, the computational results KNN algorithm will be presented and compared with those of exact method proposed by Cplex solver to validate our solution method. Then, the result of the real case study will be shown in section 4.2 by applying KNN algorithm and using Matlab software. To evaluate the validation of the solution, two methods would be tested by Matlab and Cplex software on Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.70 GHz.

4.1. Comparison between Exact method and K-Nearest Neighbor Algorithm

In this problem, two types of cost are mentioned including fixed cost for renting a vehicle and the travelling cost per distance. We defined that when each vehicle is rented, the company will pay 100,000VND for the fixed cost. It means that to reduce the total cost for transportation, the number of vehicles is used must be minimized in this problem. Besides, the traveling cost per distance is 10,000VND.

To evaluate the validation of the solution, three instances are generated as small samples for tested by exact method (using Cplex) which would result in the exact solution and algorithm (using Matlab) which is employed in the case study and would return the expected near-optimum solution. Two methods would be tested by running three generated instances from 5 nodes to 12nodes. The summary results are shown below:

Table 1. Total cost of solution from CPLEX and MATLAB

Number of Customers	Total cost - CPLEX (VND)	Total cost – KNN (VND)	Gap between KNN - CPLEX
5	4,836,000	6,077,000	0.256617039
6	5,890,000	7,727,000	0.31188455
7	6,411,000	7,268,000	0.133676494
8	6,425,000	7,745,000	0.205447471
9	6,436,000	8,512,000	0.322560597
10	8,040,000	9,766,000	0.214676617
11	9,533,000	11,301,000	0.18546103
12	11,334,000	13,385,000	0.180959944
		Average	**0.226410468**

Table 1 illustrate the objective values (total cost) in each instance from 5 to 12 nodes. As the result, K-Nearest Neighbor Algorithm always have results better than Cplex for the same nodes. The objective value is also increase followed an upward trend when increasing the number of nodes. The average gap between exact method and result of Matlab is 0.226, meaning that the K-Nearest Neighbor Algorithm is still good solution for SDVRPTW.

4.2. K-Nearest Neighbor Algorithm result apply for case study (YCH Protrade Company)

We apply K-Nearest Neighbor Algorithm model with the real dataset of the company. We used 74 trucks for delivery to 82 customers. The depot will be open 24/7. By apply K-Nearest Neighbor Algorithm, this study finds near-optimal routes so that drivers can follow the routes and finish their work. More importantly, the solution minimizes the total travel cost of the company. The summary of results when apply KNN algorithm to the real case study is shown in Table 2.

Table 2. General analysis of solution automatically achieved from Matlab

Vehicle	Total distance (km)	Total time (min)	Total delivered load per vehicle	Capacity (%)	Vehicle	Total distance (km)	Total time (min)	Total delivered load per vehicle	Capacity (%)
1	60.4	439.56	697	0.87125	38	27.8	106.68	189	0.945
2	172.6	477.12	600	1	39	29.5	107.4	185	0.925
3	137.8	435.36	600	1	40	49.3	130.8	182	0.91
4	104.2	395.04	600	1	41	57	125.88	182	0.91
5	229.8	525.48	798	0.9975	42	29	108	181	0.905
6	135.7	432.84	600	1	43	65.7	130.8	174	0.87
7	134.8	431.76	600	1	44	45.7	127.08	170	0.85
8	154.4	455.28	600	1	45	49.7	132.24	168	0.84
9	121	415.2	600	1	46	49.8	120.72	191	0.955
10	106.4	397.68	600	1	47	33.6	110.4	189	0.945
11	132.1	494.64	788	0.985	48	42.2	128.04	192	0.96
12	148.5	432.36	600	1	49	49.9	133.44	153	0.765
13	198.3	487.92	562	0.93666667	50	77.8	148.32	146	0.73
14	153.9	454.68	600	1	51	70	139.08	194	0.97
15	138.7	436.44	600	1	52	70.3	154.2	145	0.725
16	161.7	464.04	600	1	53	38.6	95.04	143	0.715
17	157.4	458.88	600	1	54	37.9	113.16	198	0.99
18	147.5	447	600	1	55	56.7	138	140	0.7
19	128.5	424.2	600	1	56	54	140.88	185	0.925
20	157.4	472.08	713	0.89125	57	40.2	104.76	136	0.68
21	103.1	393.72	600	1	58	61.2	131.28	135	0.675
22	112.9	405.48	600	1	59	43	115.8	179	0.895
23	102.7	393.24	600	1	60	85.1	160.68	193	0.965
24	123.4	418.08	600	1	61	42.1	89.4	195	0.975
25	162.2	492	727	0.90875	62	46.1	90.72	126	0.63
26	108.8	400.56	600	1	63	45.6	99.96	200	1
27	185.5	582.6	800	1	64	47.7	91.68	199	0.995
28	205.4	516.48	600	1	65	49.9	101.52	197	0.985
29	151.6	451.92	600	1	66	49.9	83.76	117	0.585
30	175.1	480.12	600	1	67	50.3	110.76	196	0.98
31	160.2	469.56	674	0.8425	68	50.1	120.12	183	0.915
32	111.4	583.68	1000	1	69	50.1	102.12	199	0.995
33	164.3	467.16	600	1	70	51.7	93.84	190	0.95
34	144.6	533.52	800	1	71	103.5	181.2	106	0.53
35	10.4	96.48	199	0.995	72	56.6	77.28	99	0.495
36	64.2	149.76	193	0.965	73	60.6	90.24	197	0.985
37	59.3	130.8	191	0.955	74	60.6	117.24	185	0.925

We we apply Nearest Neighbor Algorithm to address SDVRPTW problem for 82 customers and single depot. As as result, we used total 74 trucks to operate 152 trips in total for delivering goods to customers. The total cost for delivering for all customers is 84,750,000 VND and the total traveling distance is 16912.7 km. All customers were successfully routed within the given time windows. Based on the solution, several cost improvements were found. With the smaller number of vehicles to operate, the company can save operational cost that included fixed cost for renting vehicles, driver cost, bus maintenance cost and fuel cost. According to the data collection, we also found that this solution can save about 15% of the total cost in comparison with the existing routing and scheduling. In the reality, the company will use a single vehicle to deliver for a customer, mean that each customer owned a single vehicle for distribution process.

In this research, the objective is to minimize the total transportation cost as well as reduce the number of vehicles used which represents fixed cost for renting vehicles and fuel cost. Besides, we also consider the impacts on society, economy and environment. For economic impacts, optimizing the route for delivery helps YCH reduce transportation costs and optimize transportation productivity. Furthermore, this makes it easier for them to track and manage their distribution process. Moreover, with good routing planning for each driver, the company can ensure the productivity and well-being of its employees as they do not have to rush or work overtime to get their work done. These drivers will commit to working longer for the company. As a result, the company's revenue will not be affected by hiring new people or not meeting customer demand. For social and environmental impacts, making the most of the number of vehicles used for transportation will significantly reduce the number of vehicles, which means reducing the number of vehicles on the road. Thereby reducing risks such as traffic jams, car accidents, emit a huge amount of greenhouse gases including carbon dioxide, nitrous oxide and methane, which enormously contribute to climate change. Moreover, the few vehicles on the road, especially those with large loads, will increase the "lifetime" of the road quality. By providing optimal routes for the company, vehicle emissions may be reduced because fewer kilometers driven means fewer emissions. Hence, reducing transportation activities will contribute to environmental protection.

5. Conclusion and Recommendation

This study tries to solve these challenges through exact method and application of K-Nearest Neighbor algorithm in SDVRPTW problem. It focuses heavily on timed and split-load deliveries, which means customers will always receive their daily orders within their time windows and in case daily orders exceed capacity, load of the truck, split load was used to solve. After applying this method through Matlab software, the total distance traveled from the Nearest Neighbor algorithm has been reduced compared to the original solution. Total costs are also minimized.

However, in this report, there are still some limitations that need to be solved in the future. In this study, we only focus on homogenous fleet of trucks or trucks with the same capacity. In reality, there are some other types of trucks that can be used to deliver the order to customers based on their location and the size of order. Moreover, the travel time and travel distance in this study are estimated. However, traffic congestion and road infrastructure conditions are not considered in this research. The time windows of customers in this study have been relaxed since not many customers require a specific time period in the actual case. Moreover, the break time is ignored in this study, it is not correct in the reality.

References

Andreas Bortfeldt, Junmin Yi (2019). “The Split Delivery Vehicle Routing Problem with three-dimensional loading constraints”, European Journal of Operational Research

Binbin Panb, Zhenzhen Zhang, Andrew Lim (2020). “Multi-Trip Time-Dependent Vehicle Routing Problem with Time Windows”, European Journal of Operational Research

C. Archetti, A. Hertz, M.G. Speranza (2003). “A Tabu Search Algorithm for the Split Delivery Vehicle Routing Problem”, Les Cahiers du GERAD

Claudia Archetti, Martin W.P. Savelsbergh, M. Grazia Speranza. “An Optimization-Based Heuristic for the Split Delivery Vehicle Routing Problem”, Department of Quantitative Methods, Brescia, Italy and Georgia Institute of Technology, School of Industrial and Systems Engineering, Atlanta, U.S.A.

Damon Gulczynski, Bruce Golden, Edward Wasil (2010). “The split delivery vehicle routing problem with minimum delivery amounts”, Transportation Research Part E

E Latiffianti, N Siswanto, R A Firmandani (2017). “Split delivery vehicle routing problem with time windows: a case study”, International Conference on Industrial and System Engineering

L. Wang, J. Kinable, T. van Woensel (2020). “The fuel replenishment problem: A split-delivery multicompartment vehicle routing problem with multiple trips", Computers and Operations Research 118

LinglingDu, RuhanHe (2012). "Combining Nearest Neighbor Search with Tabu Search for Large-Scale Vehicle Routing Problem", Physics Procedia

Magnus Stalhane, Henrik Andersson, Marielle Christiansen, Jean-Franc-ois Cordeau, Guy Desaulniers (2012). "A branch-price-and-cut method for a ship routing and scheduling problem with split loads", Computers & Operations Research

Mazin AbedMohammed, Mohd Khanapi AbdGhani, Raed IbraheemHamed, Salama A. Mostafa, Dheyaa AhmedIbrahim, Humam KhaledJameel, Ahmed HamedAlallah (2017). "Solving vehicle routing problem by using improved K-nearest neighbor algorithm for best solution", Journal of Computational Science

Mingzhu Yu, Xin Jin, Zizhen Zhang, Hu Qin, Qidong Lai (2019). "The split-delivery mixed capacitated arc-routing problem: Applications and a forest-based tabu search approach", Transportation Research Part E 132

Moshe Dror, Gilbert Laporte, Pierre Trudeaubvc (1994). "Vehicle routing with split deliveries", Discrete Applied Mathematics 50

Mustafa Sahin, Gizem Çavus lar, Temel Öncan, Güvenç Sahin, Dilek Tüzün Aksu (2013). "An efficient heuristic for the Multi-vehicle One-to-one Pickup and Delivery Problem with Split Loads", Transportation Research Part C

Nicola Bianchessi, Michael Drexl, Stefan Irnicha (2019). "The Split Delivery Vehicle Routing Problem with Time Windows and Customer Inconvenience Constraints", Transportation Science

S.C. Ho, D. Haugland (2004). "A tabu search heuristic for the vehicle routing problem with time windows and split deliveries", Computers & Operations Research 31 (2004) 1947–1964

Seyed Farid Ghannadpour, Simak Noori, Reza Tavakkoli-Moghaddam, Keivan Ghoseiri (2014). "A multi-objective dynamic vehicle routing problem with fuzzy time windows: Model, solution and application", Applied Soft Computing 14

Vincent F. Yu, Nguyen Thi Thuy Linh, Phan Nguyen Ky Phuc (2020). "Vehicle Routing Problem with Perishable Products", Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan and Department of Industrial and Systems Engineering, Ho Chi Minh City International University, Vietnam

MECHANICAL PROPERTIES OF SPECIMENS 3D PRINTED WITH VIRGIN AND RECYCLED ACRYLONITRILE-BUTADIENE-STYRENE AND POLYETHYLENE TEREPHTHALATE

Thi Thanh Tam Nguyen[1], Bernd Muschard[2], Chia-Nan Wang[3]

[1]*Thi Thanh Tam Nguyen, Hong Bang International University, National Kaohsiung University of Science and Technology, tttam.nguyen@outlook.com*

[2]*Bernd Muschard, Technical University of Berlin, muschard@mf.tu-berlin.de*

[3]*Chia-Nan Wang, National Kaohsiung University of Science and Technology, cn.wang@nkust.edu.tw*

Abstract. Additive manufacturing technologies can make a significant contribution to more efficient, sustainable, and personalized products. Additive manufacturing promise over conventional production techniques diverse efficiency improvements, not only in terms of cost but also in terms of material and energy conservation, as well as improved working conditions, all reflected in the economic, environmental and social accounting. For example, a shredder and plastic extruder can turn a non-valuable object like thermoplastic domestic waste, into a valuable product like 3D printer filament. The aim of this study is the scientific analysis of 3D printer filaments, made from recycled plastic waste. For this purpose, four kinds of filament including virgin and recycled of both acrylonitrile-butadiene-styrene (ABS) and polyethylene terephthalate (PET) shall be used to produce 3D specimens by the Ultimaker 2 extended+ in order to evaluate the hardness, tensile strength, and flexibility. Beyond a critical literature review, the results of mechanical properties were revealed based on a set of experiments designed in compliance with DIN EN ISO 291 for conditioning, ISO 527 for tensile test, DIN EN ISO 178 for the flexural test, and DIN EN ISO 6507 for Vickers' hardness test. The results showed that recycling does has negative effects on quality of plastics because the polymers degrade under thermal, and mechanical impacts evidenced by the decrease of strength in tension, flexure, and indentation.

Keywords: Additive manufacturing; FDM; 3D printing; recycling; ABS; PET; 3D printer filament.

1. Introduction

The additive manufacturing (AM) has been developed in all means, and contributes their sustainable values such as material, and energy saving, reduction of failures, and reconnection of makers, and designers in three aspects, ecology, economy, and society. The fused deposition model (FDM), the most popular technique of AM, fabricates the desired products from filament loaded as a spool, the feedstock made by extrusion process. The resources of filament are usually originated from thermoplastics, a kind of polymers with linear backbone, due to the shaping ability, and toughness.

Two types of thermoplastics, ABS and PET are selected to be tested based on their suitability in making recycled filament which are recyclability, popularity in recycling, and processability. Scraps of ABS, and PET can be retrieved from the waste stream of electrical, car industry or dairy containers to be grinded, and reprocessed by mechanical recycling. The impacts on the performances after recycling can be explained by polymer degradation in which the molecules weight is reduced dramatically if the polymers are processed at high temperature for a long time.

The changes of mechanical properties are evaluated by three measurements, tensile, flexural, and hardness test. To standardize the methods, the international standards are responsible for all experimental techniques, including preparation, inspection, conditioning, and procedures.

The results of evaluation are pessimistic evidenced by the decrease of strength in tension, flexure, and indentation (figure 1). It can be concluded that the recycling does has negative effects on quality of plastics because the polymers degrade under thermal, and mechanical impacts. However, the conclusion should be regarded in certain conditions not in the whole because of uncertainties in measuring system. Therefore, the results of this research should be regarded as a reference point to conduct a comprehensive study of statistic. Moreover, the works can contribute to the further decision making of recycled plastics implementation in the additive manufacturing, and sustainability.

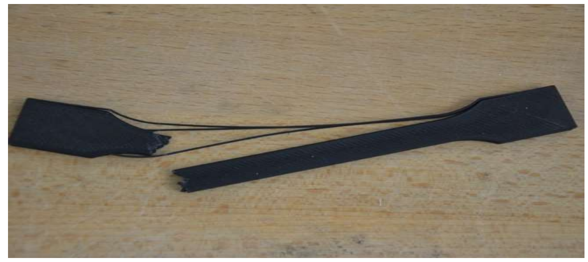


Figure 1. The brittleness and splitting of specimens made by recycled ABS after tensile test.

2. Literature Review

Additive manufacturing is the term evolved from the development of rapid prototyping (RP) understanding, a method to introduce rapidly a rough model to customers before trading (Gibson, 2010) or "A process of joining materials to make objects from 3D model data, usually layer upon layer, as opposed to subtractive manufacturing methodologies" (Committee, 2012).

Among additive technologies, FDM referring to a mechanical process to transform thermoplastics formed as filament into molded objects by extruding layers by layers with or without supporting features based on CAD designs (figure 2) (Laperrière, 2014) is one of the most popular extrusion-based printing due to cheap cost, wide range of materials, and less toxic (Freitas, 2016).

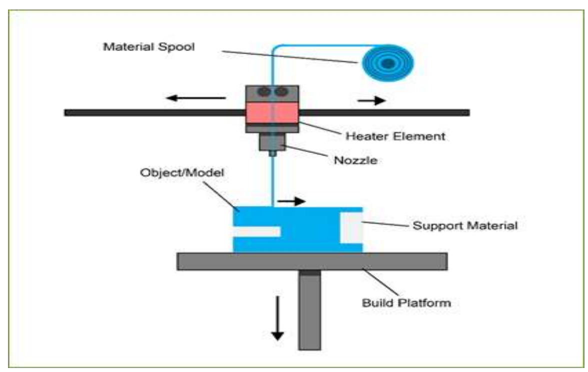


Figure 2. Materials extrusion (FDM)

In AM industry, thermoplastics, the artificial polymers made by chemical reactions or polymerization play a crucial role in material sources with nearly a half of usage according to a report of (Forster, 2015) including ABS, and PET, etc. It can be observed that the more polymers use in applications rise, the more concern in sustainability increase, respectively matters in municipal solid-waste (MSW). The majority of MSW accounts for thermoplastics with high rates in packaging (Kutz, 2007) leading to a revolution of plastics recycling. Other than ecology aspect, due to high price of virgin feedstock, the economy approach can be achieved by utilization of recycled plastics. Subsequently, making use of waste stream to supply for main stream has been researched recently. Some works can be mentioned lately like the examination of recycled (r) high density polyethylene (HDPE) filament fabrication correlating to energy usage, and time (Baechler, 2013), and several technical parameters of extrusion (Hamod, 2014) or selection of rABS for FDM based on melt flow index (Sa'ude, 2015). In 2012, Perpetual plastics project was begun to build the idea of small-scale production of recycled filament, then customers experience commercial rABS of Dimension Polymers with up to 95% of recycled rate purchased at 35$ for 750 gram of 1.775 mm filament, and from Fila-Cycle group, approximately 100% of rPLA, rPET, rABS, and rHIPS2 in 2015 or rPET of Reflow filament team just launched in 2016.

Although, the recycling may be a green solution to solve waste's crisis, it is noted that, polymers are endured thermal, and mechanical degradation relatively during the process resulting in deterioration plastics'performances (Kelen, 1983). To evaluate the mechanical properties of plastics, tensile tests may be the most significant practice to derive information about the resistance to tension commonly expressed by the tensile strength, and elongation (Brown, 2002). The test's principles practically follow the stress-strain curves (figure 3) obtaining three vital points of displacement, Yield point, peak of stress, and fractured point. Firstly, the specimen behaves elastic when elongating from l_0 to l_e at Yield stress, beginning with linear elastic behavior, proportional elongation to the load then nonlinear elastic behavior at stress's value of proportional limit. Nonetheless, the Yield strength can be difficult to find on the stress-train curve of soft, and ductile plastics due to slightly inaccuracy in slope resulting in an offset by 0.2% e. After the Yield point, although the object cannot reverse the deformation, plastic behavior, the uniform elongation still occurs until expanding to l_u at ultimate tensile strength (UTS) or tensile strength. Because UTS represents the maximum force at which the plastic can withstand, the specimen begins necking or uneven defect then ending up broken at l_f (Serope Kalpakjian, 2010).

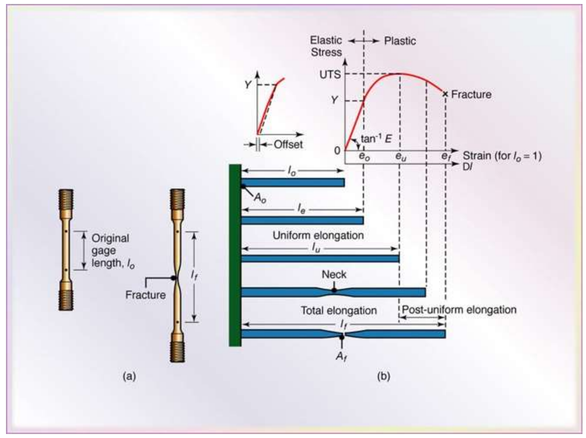


Figure 3. The principle of tension test (Serope Kalpakjian, 2010).

Similar to the tensile test, flexure test or bending test represents for a method defining the flexural strength, and deflection as a result of material's behavior in the outer layers during deformation. On the basis, the loading rod gradually contacts to a rectangular sample lying freely on two fixtures until engaging at which the force is transmitted perpendicularly at one point in three-point bending or two points in four-point bending (figure 4), and is started to count by force sensor. That procedure results in stress, and strain with two phenomena, tension on the upper surface, and compression on the lower surface (Serope Kalpakjian, 2010).

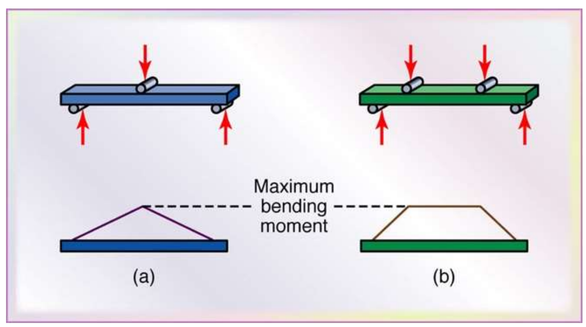


Figure 4. The three-point, and four-point bending test (Serope Kalpakjian, 2010).

Regarding to the material's strength in response to be scratched, worn, and resilient or imperishable indentation, the indicator is hardness (Serope Kalpakjian, 2010). For plastics' applicability, there are several approaches including Brinell, Rockwell, Vickers, and Knoop varied in the shape of indentation.

3. Methodology

In a common sense, the methodology is presumed as a coherent description of all gathered methods which are the appropriate means to control, and resolve the tasks. In this section, the taken actions to accomplish the requests will be introduced to the readers in organization based on material's flow as visualized in figure 5. It is obvious that the initial point must be the objects of study, the thermoplastics, and their recycles that origins are clarified in the collecting process. The following attempts are officially regulated by the international organization for standardization (ISO) which dictates the techniques in demands to implement the experiments. In order to lower the degree of errors, and wastes, the preparation in which the concept of specimen is realized, is situated before fabrication with the specific parameters by FDM model. The prints are then inspected for their conformity to proceed to conditioning phase in case of acceptable quality, otherwise they are discarded. A well-disciplined reservation with caution of temperature, humidity, and time can exceptionally provide more reliable outcomes. Last but not least, the explicit guidance of managing events, and apparatus is brought up to help the implementation at the end to be feasible.

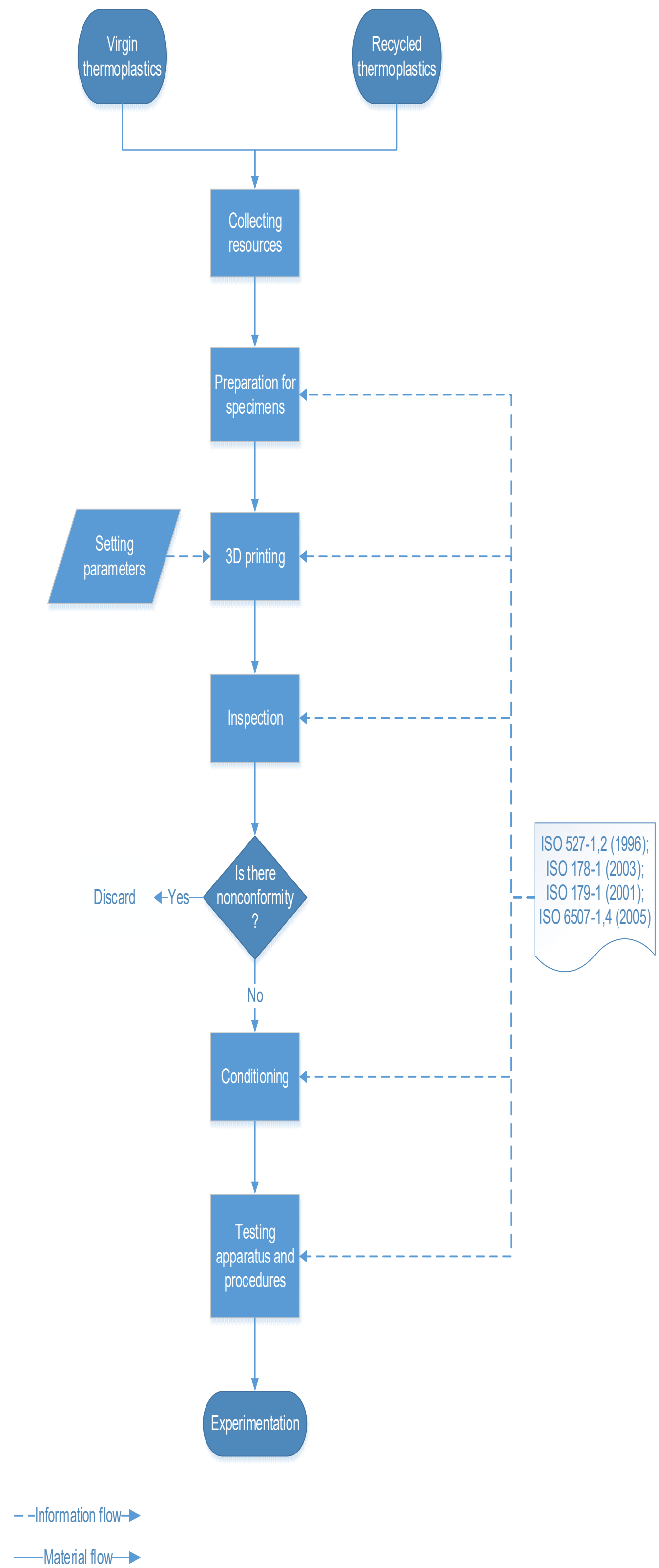


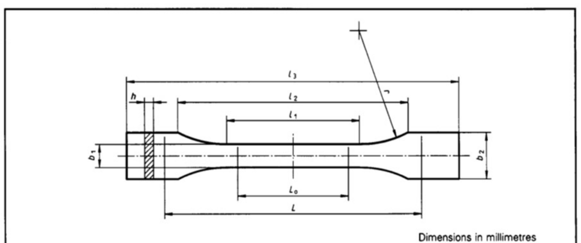
Figure 5. The research graph

3.1. Experimental techniques

3.1.1. Preparation for specimens

Tensile test

There are actually several normative references for tensile test of plastics that can be listed as ISO 527 (527-1, 1996) (527-2, 1996) or ASTM D638. However, in a report of the National Institute of Standards, and Technology (Forster, 2015), the reporter claims that in ASTM D638 (D638, 2002), premature break outside of gage length can happen due to the specifications of shape, and dimensions. As a matter of fact, ISO 527 is chosen to specify type of specimens for directly-moulded multipurpose as 1A (527-2, 1996) (figure 6), and number of objects as 5 respecting to probability of 95%in measuring (527-1, 1996).



Dimensions in millimetres

	Specimen type	1A	1B
l_3	Overall length	≥ 150 1)	
l_1	Length of narrow parallel-sided portion	80 ± 2	60,0 ± 0,5
r	Radius	20 to 25	≥ 60 2)
l_2	Distance between broad parallel-sided portions	104 to 113 3)	106 to 120 3)
b_2	Width at ends	20,0 ± 0,2	
b_1	Width of narrow portion	10,0 ± 0,2	
h	Preferred thickness	4,0 ± 0,2	
L_0	Gauge length	50,0 ± 0,5	
L	Initial distance between grips	115 ± 1	$l_2{}^{+5}_{0}$

Figure 6. Tensile specimen type 1A, and 1B (527-2, 1996).

Flexural test

In the case of testing flexure, either ASTM D790 (D790, 2000) or DIN EN ISO 178 (178, 2003) can be applicable as claimed by (Forster, 2015). In the research, ISO 178 is used to standardize geometry, and quantity of specimens in order to reveal flexural strength, and elongation during three-point bending process. It is required 5 rectangle pieces in the design of flexural measurement (figure 7).

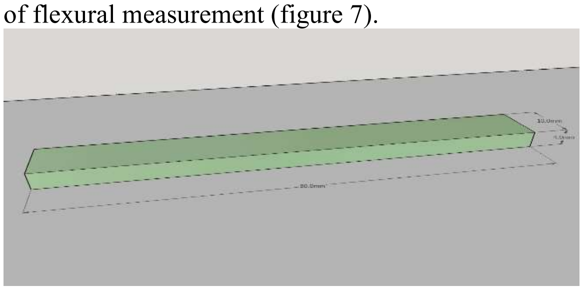


Figure 7. Design of flexural testing object.

Hardness test

Generally speaking, to control ABS, and PET's hardness, Rockwell or ball indentation method is more preferable in many documents (E.Mark, 1999). Nonetheless, the Vickers practice, popular for metallic materials, is more, and more reported in several polymer testing references despite the fact that it has not been standardized yet (Brown R. , 2002). The designation of testing objects are principally guided by ISO 6507-1 (6507-1, 2006) including some steps, searching the range of hardness value, identifying the kind of test, and calculating the dimensions.

The Vickers hardness value (HV) of ABS can be found, ranged from 5.6 to 15.3 HV, in case of difficult finding like HV of PET, it may be calculated from Yield strength following to the equation 3-1 which as Zhang (Zhang, 2011) says in the author's journal, approximately proves the correlation of HV, and strength by empiricism:

$$\boldsymbol{HV \approx 3 \times \sigma_Y} \quad (1)$$

If HV is expressed in kg.f/mm2, and σ_Y expressed in MPa then equation (1) leads to equation (2):

$$\boldsymbol{HV \approx 3 \times 0.102 \times \sigma_Y} \quad (2)$$

Where 0.102 is a conversion factor from MPa to kg.f/mm2.

Based on equation (2), and σ_Y of PET= 56.5 – 62.3 MPa, the HV of PET is formed below:

$$HV_{PET} \approx 3 \times 0.102 \times \{56.5 - 62.3\}$$
$$\approx \{16.95 - 18.69\}\ \text{HV}$$

From the results above, the limit of hardness symbol, and test force, also mean of the indenter's diagonal length ($\bar{d}$) can be identified in ISO 6507-4 (6507-4, 2005). It can be argued on the data of ISO 6507-4 (6507-4, 2005) that both Vickers microhardness test, and low-force Vickers hardness test are feasible to acquire the expected HV, however, evidence shows decreasing test force, and $\bar{d}$ may worsen the measurements. Hence, the tests are proceeded with HV 0.5, test force of 4.903 N in the range of low-force type.

The minimum of specimen' thickness is then designed by at least 1.5 times of largest $\bar{d}$ (6507-1, 2006) which is 0.222 mm corresponding to 18.8 HV the maximum possibly measured HV (6507-4, 2005), thus the lower limit is nearly equal to 0.333 mm. The specimen is eventually made up as round model aiming to minimize the material with the thickness of

6 mm, and diameter of 26 mm (figure 8) to serve two essential purposes, succeed the indentation, and perform multiple marks. In contrast to the previous techniques, Vickers test's quality is assured at 95% by 5 contacting times on 1 sample (6507-1, 2006).

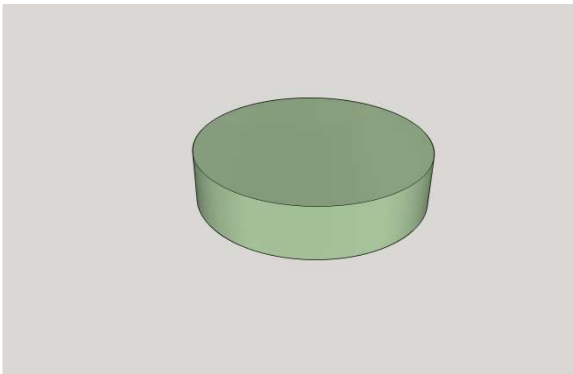


Figure 8. Design of Vickers testing object

3.1.2. 3D printing

All the test pieces are fabricated by Ultimaker 2 extended+, a FFF machine of the Ultimaker Ltd. Company, also a Cartesian printer which is framed with three axes (Horvath, 2015).

Each type of thermoplastics is replicated respecting to its particular setting to ensure the quality as presented in table 1, and 2. It cannot be ignored the fact that layers height, and percentages of infill may be significant influences to the mechanical performances. Hence, the fill density of 100%, and layers height of 0.25 mm as investigated seem to give the optimal properties.

Table 1. 3D printing parameters of virgin plastics

Parameters Material	ABS	PET
Nozzle Temperature	260 ^{0}C	230 ^{0}C
Buildplate temperature	80 ^{0}C	60 ^{0}C
Layers height	0.25 mm	0.25 mm
Fill density	100%	100%
Print speed	40mm/s	40mm/s
Nozzle size	0.6 mm	0.6 mm
Fan speed	0%	40%

Table 2. 3D printing parameters of recycled plastics

Parameters Material	rABS	rPET
Nozzle Temperature	240 ^{0}C	230 ^{0}C
Buildplate temperature	100 ^{0}C	60 ^{0}C
Layers height	0.25 mm	0.25 mm
Fill density	100%	100%
Print speed	40mm/s	40mm/s
Nozzle size	0.6 mm	0.6 mm
Fan speed	0%	10%

3.1.3. Inspection

After printing, the specimens are inspected regarding to their surfaces, sizes, and shapes by the visual examination first, and then by micrometer caliper Holex® version 41 2808, resolution 0.01mm, error limit 0.03. The visual check is targeted to discover possible nonconformity in the outer faces, and outline whereas the caliper is meant for the evaluation of uniformity (527-1, 1996). It is true to say that meeting requirements of norms is the specimen inspection's goal so the aims of each type of test will be distinct. For example, either flexural or tensile test pay more attention in the control of dimensions, especially the width, and thickness due to their relevance to the calculation later, in the centre of the test pieces where the deformation highly occurs.

3.1.4. Conditioning

The conditioning is a mean of restraints on the surrounding conditions which are regularly planned for two periods of test, the preparation in which the specimens are prevented from the degradation, and preserved to approach material's equilibrium, the execution in which the atmosphere of laboratory is adjusted during the test (Brown R. , 2002).

There is a fact that plastics is likely to be very sensitive to ambient conditions in such a manner that the air temperature, humidity, and storage's time is seriously restricted. The standard atmosphere in conformity with DIN EN ISO 291 (291, 1997) is practically preferred to acquire as a pair of 23±20C, and 50±10% which are specified for both conditioning, and test atmosphere. The minimum storing duration of 40 hours, a significant factor to balance the material's state, is referenced for the thick pieces over 7 mm (D618, 2005).

In the research, the specimens are conditioned in air-conditioned room to sustain the conditioning temperature, and in the vacuumed plastic bags with silica gel to probably stay out of moisturizing meanwhile the test atmosphere is taken care by the utility of air-conditioned chamber only.

4. Results

4.1. Tensile test

In this paper, tensile properties are evaluated in a way of involving only engineering values of extreme tensile stress, and strain. The equation (3), and (4) reflect the calculation of ultimate tensile stress that outputs are detailed in table 3 for ABS samples, and table 4 for PET samples. The estimation of strain is split into two definitions because it depends on which the original length is considered, if counting gage length then defining as strain, else nominal strain if grips separation is counted (equation (5)). The results of nominal strain are also recorded in table 3, and table 4, and rounded to two decimal places whereas three decimal places in tensile stresses. In a graphic view, the line charts below reveal the stress-strain trend of thermoplastics in virgin (figure 9, 11), and reprocessed state (figure 10, 12).

$$\sigma_M = \frac{F_M}{A} \qquad (3)$$

where

σ_M: The maximum of engineering tensile stress in question, MPa;

F_M: The maximum of measured force, N;

A: The cross-sectional area of dog-bone piece, mm2.

$$A = b_1 \times h \qquad (4)$$

where

b_1: The width of narrow part, mm;

h: The thickness of specimen, mm.

$$\varepsilon_{tB} = 100 \times \frac{\Delta L}{L} \qquad (5)$$

where

ε_{tB}: The nominal tensile strain at break in question, %;

ΔL: The increase of grips distance until fracture, mm;

L: The initial grip to grip separation, mm.

Table 3. Tensile properties of virgin, and recycled ABS

Test pieces	Tensile strength (MPa)		Tensile nominal strain at break (%)	
	ABS	**rABS**	**ABS**	**rABS**
1	29.923	22.652	4.18%	2.59%
2	34.037	25.252	5.08%	2.94%
3	35.476	25.479	5.32%	2.91%
4	33.928	25.051	4.86%	2.97%
5	35.204	25.044	5.63%	2.94%
Mean	33.713	24.696	5.02%	2.87%

Table 4. Tensile properties of virgin, and recycled PET

Test pieces	Tensile strength (MPa)		Tensile nominal strain at break (%)	
	PET	**rPET**	**PET**	**rPET**
1	64.878	38.964	6.45%	5.74%
2	65.495	40.587	5.29%	5.67%
3	65.735	34.024	5.26%	4.85%
4	66.184	32.931	5.14%	4.68%
5	67.421	30.682	5.48%	5.29%
Mean	65.943	35.437	5.26%	5.52%

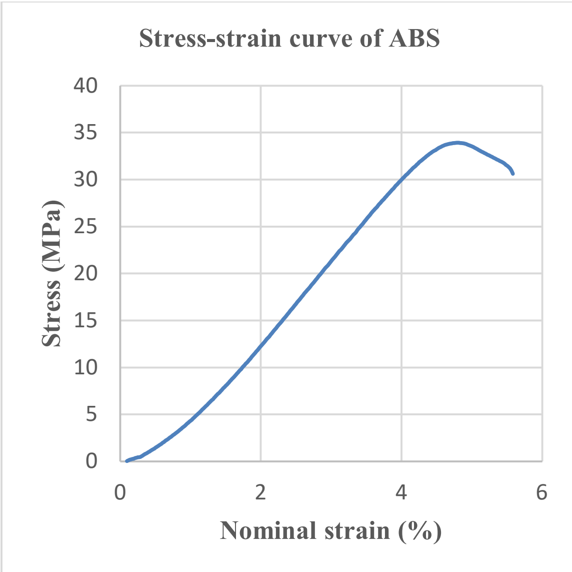


Figure 9. Tensile behavior of virgin ABS

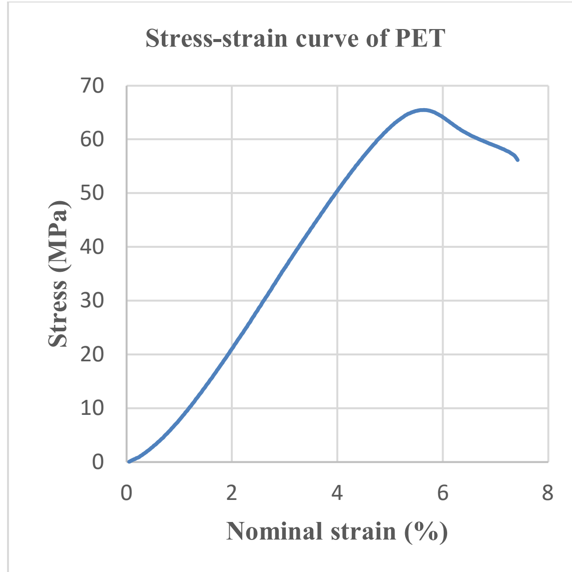


Figure 11. Tensile behavior of PET

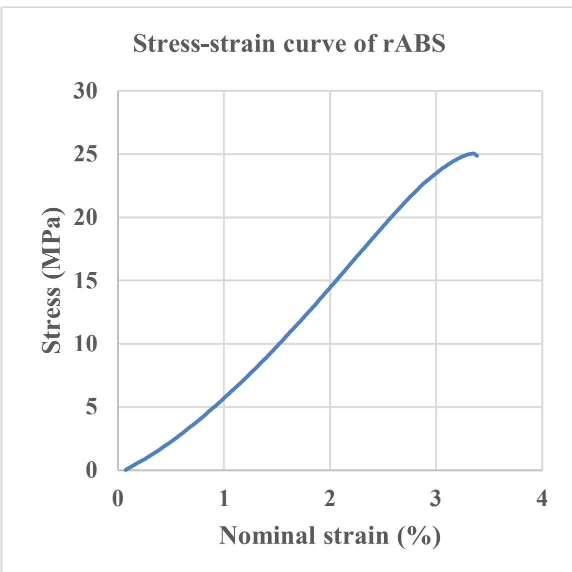


Figure 10. Tensile behavior of recycled ABS

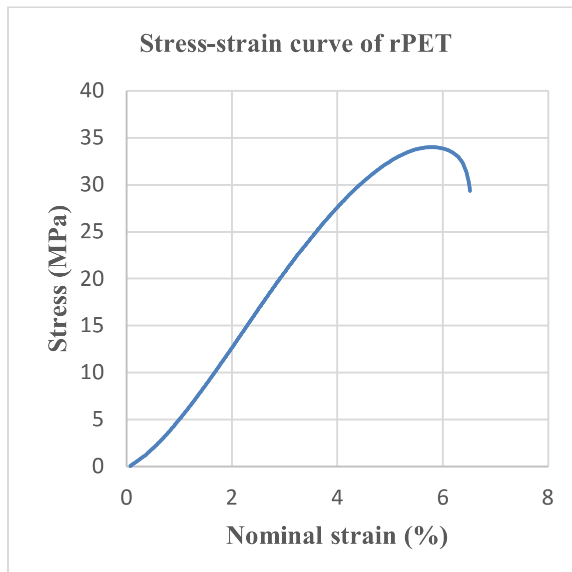


Figure 12. Tensile behavior of recycled PET

4.2. Flexural test

The most vital properties of flexure, flexural strength which the maximum force is accounted for, shows the ultimate stress of material, corresponded to formula below:

$$\sigma_{fM} = \frac{3F_M L}{2bh^2} \qquad (6)$$

where
σ_{fM}: The maximum flexural stress in question, MPa;

F_M: The maximum force in measuring, N;

L: The span, mm;

b: The test beam's width, mm;

h: The test beam's thickness, mm.

The ultimate flexural strength of five rectangular beams of each type of plastics with their average values are calculated, and accurate to three significant figures (178, 2003) as can be seen in table 5, and table 6.

Table 5. Flexural properties of virgin, and recycled ABS

Test beams	Flexural strength (MPa)	
	ABS	rABS
1	52.039	44.779
2	52.088	42.162
3	52.118	42.239
4	50.798	42.874
5	52.582	40.983
Mean	51.925	42.607

Table 6. Flexural properties of virgin, and recycled PET

Test beams	Flexural strength (MPa)	
	PET	rPET
1	98.769	54.972
2	94.530	50.617
3	96.737	54.637
4	95.263	51.930
5	94.603	49.405
Mean	95.980	52.312

4.3. Hardness test

The calculation of hardness values involves with the mean of the diagonals' length represented for surface area of indentation, and relates to the below equation:

$$HV = 0.102 \times \frac{2 \times F \times sin\frac{136^\circ}{2}}{\bar{d}^2} \approx 0.1891 \times \frac{F}{\bar{d}^2} \quad (7)$$

Where

HV: Vickers hardness value;

F: Test force, N;

$\propto$ = 1360: the angle between the opposite faces at the vertex of the pyramid;

$\bar{d}$: The mean of the two diagonal lengths d1, d2.

Based on ISO 6507-1 (6507-1, 2006), 5 hardness assessments of each plastics are obtained, and accounted for their arithmetic mean which is only taken two decimal numbers (table 7, 8).

Table 7. Vickers hardness values of virgin, and recycled ABS

Indentations	Vickers hardness value	
	ABS	rABS
1	9	9
2	8	8
3	8	10
4	9	10
5	9	11
Mean	8.60	9.60

Table 8. Vickers hardness values of virgin, and recycled PET

Indentations	Vickers hardness value	
	PET	rPET
1	18	11
2	16	9
3	16	11
4	16	11
5	15	10
Mean	16.20	10.40

5. Discussion

On the whole, both ABS, and PET witness the sharp decline of tensile properties when comparing to reprocessed samples. Even though, the loss of tensile strength, and nominal strain in rPET is nearly a half of its virgin meanwhile rABS is worse than its origin about one-third of strength, and resemble most of strain percentages. It is undeniable that the reduction of strength, and elongation have impacts on material's performance that are more brittle, and less tough. The changes are visualized in the stress-strain trend lines which illustrate the recovered plastics quickly break after reaching the peak of stress (figure 4-2, 4-4) in contrast to their origin (figure 4-1, 4-3).

Likewise, the ability to withstand the pressure in outer layers drops substantially in recycled thermoplastics either rPET or rABS. There is also a greater fall of flexural strength in rPET than rABS which is approximately fifty percentages. Also, it is clear to see that PET's durability is fairly twice as much as ABS's durability. Because of high ductility, only non-recycled PET stays bending without cracking until sliding out of its fixation.

It can be interpreted that the differences in strength to endure the indentation between ABS, and its recyclate which is not enough of 1 HV point, is less dramatic than PET's situation which is approximately 6 HV point. Nonetheless, it cannot be denied that the recycling may weaken the hardness of both ABS, and PET.

Despite the efforts to assure the reliability of results, the presence of variability associated with the resources, and measuring system may still exist. Frankly, it is hard to say about how much quality of regenerated plastics is certified in the way that the manufacturer claims bluntly about the weight of recycled content, the data of recycling process, especially additives, fillers, contaminants, etc. Not only are variables in the materials but also in the inconsistent replication evidenced by the deviation of weight.

In contradiction to the belief in the equivalent of numerous compliances, the bias seems not to be out of concerns. That can be a reason to limit the results when testing procedures thorough the study are vastly in obedience to the international organization for standardization (ISO). What is more, either assigning the works to any norms, in fact, the causes of uncertainties, equipment's calibration, repeatability, resolution, atmosphere, and alignment errors may not be deniable.

6. Conclusion

To conclude, so far the data of examination has unveiled the fact that neither ABS nor PET can reserve most of their mechanical performances, notably the tensile, and flexural characteristics as well as the hardness after reprocessing. The explanations may be leant on the matter that polymers degrade under thermal, and mechanical impacts. While that may be true to some extent the expectancy is not met regarding optimistic results which reports that there are almost no changes in the mechanical properties up to five cycles of regeneration. On contrast, it will be unfair to compare among the events because of the research's gap, wherefore any conclusions wrapped up in this research should be subjected to no more than a reference point in certain situation.

Reference

ASTM D618 (2005), *Standard Practice for Conditioning Plastics for Testing.*

ASTM D638 (2002), *Standard Test Method for Tensile Properties of Plastics.*

ASTM D790 (2000), *Standard Test Methods for Flexural Properties of Unreinforced, and Reinforced Plastics, and Electrical Insulating Materials.*

Baechler, C., DeVuono, M., and Pearce, J.M. (2013), Distributed recycling of waste polymer into RepRap feedstock, *Rapid Prototyping Journal*, Vol. **19** No. 2, pp. 118–125.

Brown, R. (2002), *Handbook of Polymer Testing Short-term Mechanical Tests*, Rapra Technology Limited, UK.

DIN EN ISO 178 (2003), *Determination of Flexural Properties of Plastics.*

DIN EN ISO 291 (1997), *Standard Atmospheres for Conditioning, and Testing of Plastics.*

DIN EN ISO 527-1 (1996), *General Principles for The Determination of Tensile Properties of Plastics.*

DIN EN ISO 527-2 (1996), *Determination of Tensile Properties of Plastics.*

DIN EN ISO 6507-1 (2006), *Metallic Materials – Vickers Hardness Test – Part 1: Test Method (ISO 6507-1:2005).*

F42 Committee (2012), Terminology for additive manufacturing technologies, *ASTM International*, West Conshohocken, PA.

Forster, A.M. (2015), *Materials Testing Standards for Additive Manufacturing of Polymer Materials: State of the Art, and Standards Applicability,* National Institute of Standards, and Technology.

Freitas, D., Almeida, H.A., Bártolo, H., and Bártolo, P.J. (2016), Sustainability in extrusion-based additive manufacturing technologies, *Progress in Additive Manufacturing*, Vol. **1** No. 1-2, pp. 65–78.

Gibson, I., Rosen, D.W., and Stucker, B. (2010), *Additive Manufacturing Technologies*, Springer US, Boston, MA.

Hamod, H. (2014), Suitability of recycled HDPE for 3D printing filament, *Plastics Technology*, Arcada University of Applied Science, 12/2014.

Horvath, J.C., and Cameron, R. (2015), *3D Printing with MatterControl, Technology in Action*, Apress; Distributed to the book trade worldwide by Springer, [Berkeley, CA], New York, NY.

ISO 6507-4 (2005), *Metallic Materials — Vickers Hardness Test —Part 4: Tables of Hardness Values.*

James E.Mark (Ed.) (1999), *Polymer Data Handbook*, Oxford University Press.

Kelen, T. (1983), *Polymer Degradation*, 1. print, Van Nostrand, New York [u.a.].

Kutz, M. (2007), *Environmentally Conscious Materials, and Chemicals Processing*, Wiley series in environmentally conscious engineering, John Wiley & Sons, Hoboken, N.J.

Laperrière, L., and Reinhart, G. (2014), *CIRP Encyclopedia of Production Engineering*, Springer Berlin Heidelberg, Berlin, Heidelberg.

Sa'ude, N., Kamarudin, K., Ibrahim, M., and Ibrahim, M.H.I. (2015), Melt flow index of recycle ABS for fused deposition modeling (FDM) filament, *Applied Mechanics, and Materials,* 773-774, pp. 3–7.

Serope Kalpakjian, S.S. (2010), *Manufacturing Engineering, And Technology,* 6th ed., Pearson.

Zhang, P., Li, S.X., and Zhang, Z.F. (2011), General relationship between strength, and hardness, *Materials Science, and Engineering*: A, Vol. **529**, pp. 62–73.

QUALITY MANAGEMENT SYSTEM IN GARMENT INDUSTRY A CASE OF THUAN PHUONG GROUP

Pham Cao Van[1], Bui Dinh Tan[2] and Vo Trong Cang[3]

[1] Binh Duong University, Binh Duong, Viet Nam, pcvan@bdu.edu.vn

[2] QA Manager, Thuan Phuong Group, tan.bd@tpgroup.com

[3] Binh Duong University, Binh Duong, Viet Nam, vtcang@bdu.edu.vn

Abstract. Product quality is an especially important factor for garment manufacturing enterprises. Companies always want to create products of the highest quality, with the lowest defect rate and rework costs to be able to compete better and increase customer satisfaction. However, quality management is a difficult problem that not all companies can solve well. Nowadays, industry 4.0 has gradually developed and applied to many fields in life, production and services. Quality management system (QMS) based on industry 4.0 platform is an effective solution for quality management in garment enterprises. QMS helps managers to have an accurate, immediate view of the quality situation, easily recognize problems and provide correct solutions. The article will describe the process of designing and implementing a quality management system (QMS) at Thuan Phuong Group. After implementing the software, the defect rate decreasing from 2.3% to 1.8% and saving 20% working time of QC department. Hopefully this article will help businesses have an overview of a quality management software system in the garment industry.

Keywords: Quality Management System (QMS); Industry 4.0; Garment.

1. Introduction

Textile and garment has always been a pioneer industry in the economic development strategy, bringing great profit for our country. However, with the economic opening-up and globalization trend, the competition in the market has intensified not only domestic enterprises but also foreign enterprises. Under the circumstances, improving product quality to satisfy customer and market demand is a big factor determining the maintenance and development of the business. Therefore, product quality is always a top priority at all times.

Nowadays, with the continuous development of technology, quality management based on industry 4.0 is a breakthrough solution that helps enterprises manage product quality better than traditional methods. Through a convenient and instant system from production, quality managers can be able to aware of the quality situation of the factory at any time. In addition to a quick and accurate reporting system, it helps to solve quality problems quickly, without the occurrence of mass errors that are not detected in time. Because of these great benefits, the research and application of industry 4.0 in quality management has been gradually studied and put into practice by enterprises.

2. Theory

Quality has many different definitions: Based on the customer's point of view, Edwards (1968) defines that quality is the ability to satisfy customer needs; From the point of view of manufacturing, quality is the degree to which a particular product conforms to a design or specification (Gilmore, 1974); According to Broh (1982), quality is the best level with reasonable price and variable control at reasonable cost.

Research conducted at Thuan Phương Group. The company manufactures clothes, assigns export services to Haddad, Target...Therefore, the theoretical basis of research will be based on a number of standards on the garment industry and quality requirements of customers.

Quality management in the garment industry will include both material management and production management. Therefore, the study follows the direction based on the theoretical basis of the quality of these two groups. Raw Material Quality Management and Product Quality Management will be described in the following two sections.

2.1. Raw Material Quality Management

Checking the quality of fabrics before putting them into production is very important, to prevent defective products due to raw materials. The fabric quality control standard commonly applied by garment companies is the 4-point system of fabric inspection, a standard established according to the ASTM D5430-07 standard (2011). This is an important standard in the assessment, resolution and complaints in material quality management.

The 4-point system is based on the following control principles:

- All errors are assigned points.
- All extruded yarns (thin or thick) are wrong (unless the fabric is like that).
- Multiple continuation errors per Yard (=0.9144 m) are counted as 4 points.
- Each flaw, tear or broken yarn no matter how big or small is considered a serious error and is counted as 4 points.

How to calculate points in the 4-Point System:

- The length of defect from 0 inch to 3 inch (7,6 cm): 1 point;
- The length of defect from 3 inch to inch (15,2 cm): 2 point;
- The length of defect from 6 inch to 9 inch (22.8 cm): 3 point;
- Defect length greater than 9 inches (22,8 cm): 4 point.

Each yard of length must not count more than 4 points. Both horizontal and vertical errors have the same point.

Acceptance criteria of the 4-point system:

Each fabric roll must not exceed 20 points / 100 square yards, if the fabric roll's score is more than 20 points / 100 square yards, the roll is not qualified.

The formula for calculating the 4 Points System for 1 Fabric Roll is as follows:

$$\frac{Point\ in\ 1\ fabric\ roll}{Fabric\ roll\ length\ (Actual)(yard)} x \frac{36}{Width\ (Actual)(inch)} x100 = point - 100\ yard^2 , \quad (1)$$

The calculation formula of the 4-point system for a lot of fabric is as follows:

$$\frac{Total\ points}{Fabric\ roll\ length\ (Actual)(yard)} x \frac{36}{Width\ (Actual)(inch)} x100 = point - 100\ yard^2 , \quad (2)$$

Common fabric defects include:

- About yarn: yarn extruding on the surface of the fabric, yarn error, uneven yarn, other yarn mixed in;
- Structural errors: yarns, holes, fabric with ripples, wrinkles, folds;
- Dyeing error: staining spots or streaks, staining;
- Other faults: smelly fabric, rotten, moldy fabric, termite, glued or other insect damaged fabric;
- Unacceptable slanting fabric with tolerances for each type of fabric as follows: woven fabric: maximum allowable 2% of fabric size; Knitted fabric: maximum allowed 4% of fabric.

2.2. Product Quality Management

This is the main stage of the machining process, where defective products form. Therefore, if this stage does not establish a strict inspection process, it will increase production costs and recycling costs, leading to an increase in product costs.

Purpose: Early detection of errors in production caused by consciousness, human level, equipment and other technological factors to take timely remedial measures.

- Sewing workers: when receiving products from the previous stage, they must check 100% and After sewing, they must check 100% products of their own stages
- QC staff (Quality Control): check 100% products of the sewing line.

Inspection tools include: Inspection forms, operational documents, sampling plans, standard samples (for visual QC staff), tape measure (must be standard) which can be manufactured from glass wire , metal, descriptions of measurements, measurement locations for complex operations, quality assessment reports, perform the test in a well-lit place.

3. Methodology

Methodology of the research follows steps (table 1) :

Table 1. Article research process

Num.no	Process	Implementation
1	Research Theoretical Basis	Learn and select: + Published documents or Published researchs + Several articles in other scientific journals + Several best-documented on website are related the research content.
2	Research actual situation of management at the Company	Go through: + View company documents + Ask and review information from Quality department. + Observed the practice in company.
3	Design software system.	Base on the requirement of the factory, write the requirement and create a demo in excel software
4	Write the software	Vocational IT Develop
5	Test the software	Base on funtions have been required to implement testing software
6	Training and implement	Training to operate and implementing the software system into production

4. Write Business Requirement Document (BRD)

4.1. Raw Material Quality Management System

Method to receive input data:

Currently, the factory is using an RFID system (Radio Frequency Identification) to manage the fabric roll inventory, each fabric has been fitted with a chip that carries the information of that fabric (code, color, size, supplier, lot, etc.). Therefore, when designing Raw Material Quality Management System, the research used this chip through a chip reader to transmit data to Raw Material Quality Management System (Figure 1).

Figure 1. Input device of Raw Material Quality Management System.

Import data:

Through understanding the actual needs of users to enter what data and what reports this data will be used for, the research has omitted the data that can be obtained directly from the SAP production management system (System Application Programming) and retain the data to be entered. Then, arrange these data into 2 parts as follows.

- Fabric quality, including defect and defect points, including:
  - Defect: include common defects when inspect fabrics: yarn errors, color splashes, spots, print loss, tears, oil stains and other defects;
  - Defect points: including 4 points according to the 4-point system;
- Fabric parameters: including entering the length of the fabric roll, the width of the fabric roll (first, middle, and end of the fabric roll).

Export report:

Based on input data, available data in production management software SAP (System Application Programming) and requirements of the factory quality management department, the research has built the following reports:

- Fabric quality report, view from date to date: help managers know the fabric quality status of the factory on a day-to-day basis;
- Detailed quality inspection report: is a type of report requested from customers in the 4-point inspection system, giving detailed information of the tested fabric roll and the score of the fabric roll;
- Total quality inspection report: is a type of report requested from the customer in the 4-point inspection system, giving the aggregate information and the score of the lot fabric roll;

- Detailed supplier quality report: helps managers know the fabric quality situation of each supplier, each of lot fabric;
- Supplier quality report overview: helps managers compare fabric quality between suppliers, thereby comparing, evaluating and choosing the best supplier;
- Report on working efficiency from day to day: evaluate the work results of QC staff based on the quantity and quality of fabric rolls inspect during the day;
- Online QCM user report: helps managers know how many employees are currently inspected fabric, how many fabric inspection machines are in operation;
- Report of unfinished fabric roll check from date to date: in the process of inspect fabric, there are unfinished fabric roll, if there is a problem, it must be stopped. This report helps to display detailed information about these fabric roll.

Hardware device:

After considering the features of the software, the research has selected a hardware device that interacts between the user and the software, which is a touch screen. Thereby, the user's operation is faster, the user is more used to using it and avoids errors during use.

4.2. Product Quality Management System

Method to receive input data:

The planning department has allocated the production plan on the production management software (SAP). Therefore, the research has linked SAP and Product Quality Management System software so that the software will automatically receive input data.

Import data:

Through understanding the actual needs of data needed for reporting, the research has designed the data entry part for the Product Quality Management System including 2 parts:

- List of sewing process: taken in SAP software data system;
- The list of defect that occur frequently at the sewing line includes: broken thread, loose thread, not hugging elastic, missing thread, pleated pleats, button bugs, oil/rust, different color, sewing technique, skipping, stitching only mismatched, drooping eyelids, drooping eyelids, uneven, asymmetrical, dirty, defective materials, pending.

***Export report*:**

Based on the input data, the data available in the production management software (SAP) and the requirements of the factory quality management department, the research has built the following reports:

- Report to user QCM online: display quantity and information of QC staff working at the sewing line;
- Pre-wash parameter inspection report by date, style: helps managers monitor and ensure pre-wash parameters meet customer standards;
- The first product inspection report is viewed by date, style: when a new product style is uploaded, through this report, we know the quality problems that need to be resolved;
- The quality inspection report of QA (Quality Assurance) is viewed by date, style: the report monitors the overall quality of the factory, each sewing line, each process and each type of defect.

Hardware device:

After considering the features of the software, the research has selected a hardware device that interacts between the user and the software, which is a touch screen. Thereby, the user's operation is faster, the user is more used to using it and avoids errors during use.

5. Write Software, Implementation and Results

As analyzed, the program is built on the C# language and is divided into two software modules: Raw Material Quality Management System and Product Quality Management System (described in the following two sections).

5.1. Raw Material Quality Management System

After receiving business requirement document, the IT department wrote the software. The data entry interface (Figure 2) and the reporting system (Figure 3) are generally described as follows:

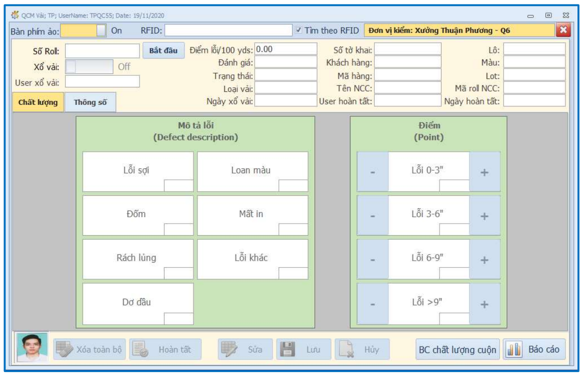


Figure 2. Data input interface of Raw Material Quality Management System

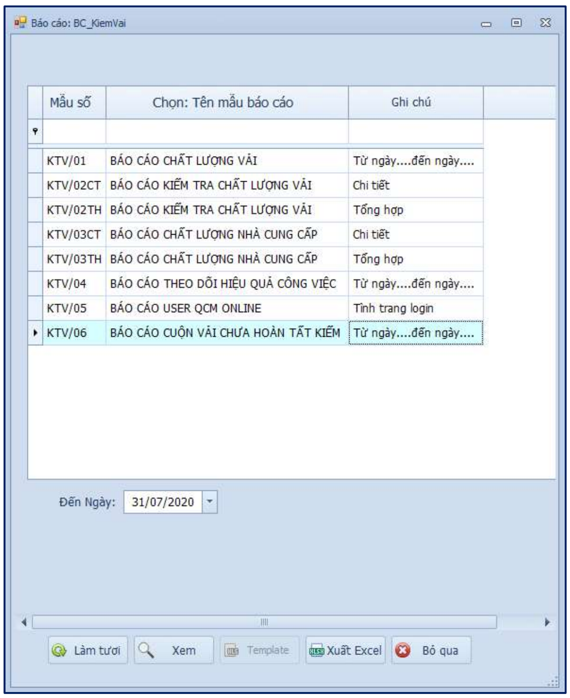


Figure 3. Report interface of Raw Material Quality Management System

Software operation process:

Step 1: QC staff remove the chip on the cloth and scan it into the card reader.

Step 2: After the fabric roll information is displayed on the screen, the QC staff presses the "Start" button to start the fabric inspection process (Figure 4).

Figure 4: Step 2 in the operating procedure of Raw Material Quality Management System

Step 3: During fabric inspection, must strictly follow the 4-point fabric inspection regulations. If fabric roll has defect, import the defect and the number of defect points on touch screen.

Step 4: During the inspection of fabric roll at the top, middle and bottom, QC staff measure the width fabric and fill it in the "Parameters" Tab. At the end of the fabric inspect process, measure the length of the fabric roll displayed on the fabric inspect machine and fill in the length information in the "Parameters" Tab (Figure 5).

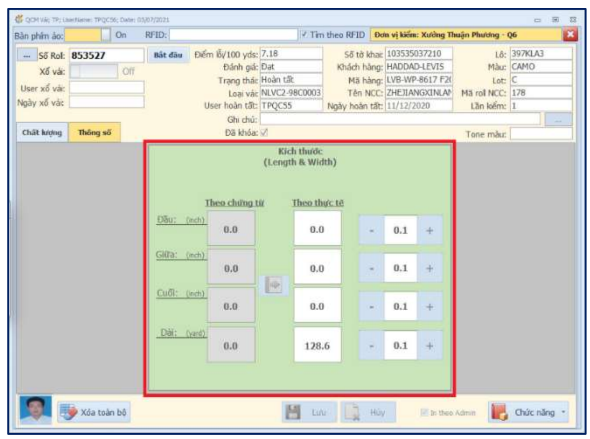


Figure 5: Step 4 in the operating procedure of Raw Material Quality Management System

Step 5: When the fabric inspection is finished, the staff presses the "Finish" button on the screen (Figure 6) and continues from step 1 to step 5 for other fabric rolls.

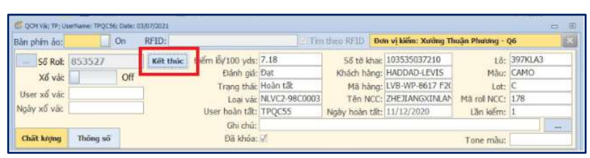


Figure 6: Step 5 in the operating procedure of Raw Material Quality Management System

After the software is written, tested and completed, Raw Material Quality Management System has been put into operation. The software has helped QC department to no longer take notes and calculate manually, but instead has been automated 100%. Thereby, a 20% reduction in the total daily working time of employees (Figure 7). Besides, fabric inspection reports are always updated continuously and accurately to help managers react promptly when fabric quality problems occur.

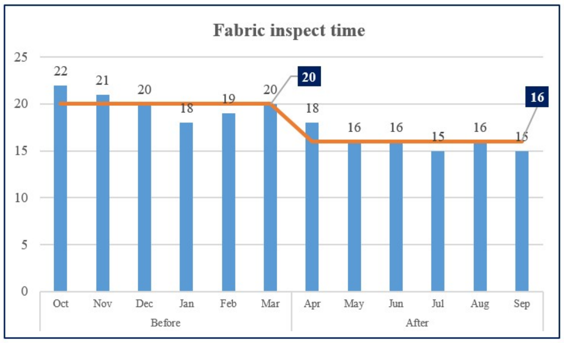


Figure 7. Fabric inspect time before and after software implementation

5.2. Product Quality Management System

Similar to Raw Material Quality Management System, after receiving business requirement document from research, the IT department had written Product Quality Management System software. The interface for data entry (Figure 8) and reporting (Figure 9) is as follows:

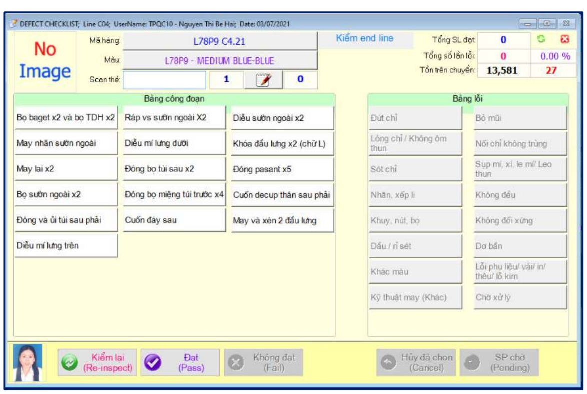


Figure 8. Data input interface of Product Quality Management System

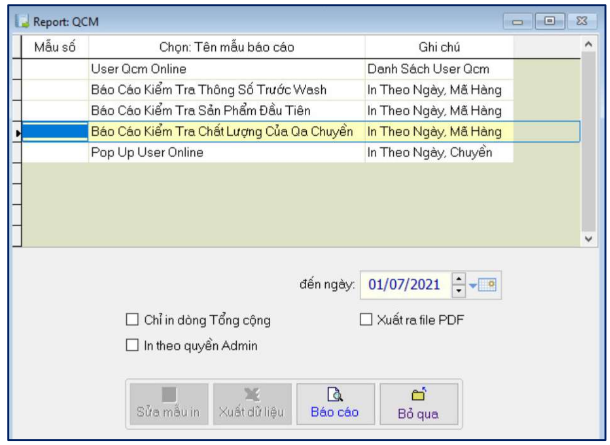


Figure 9. Report interface of Product Quality Management System

Software operation process:

Step 1: The QC staff of the sewing line selects the product style and color of the product being checked (Figure 10).

Figure 10. Step 1 in the operating procedure of Product Quality Management System

Step 2: During the inspection, if the product is defective, click on the process name and defect on the touch screen (Figure 11).

Step 3: When the QA department checks the product parameters according to AQL 2.5 standard, select the function "check parameters before wash" and fill in the parameter information (Figure 12).

Product Quality Management System software has been applied at the sewing lines of Thuan Phuong Group. After implementing the software for 6 months (from April 2020 to October 2020). The software has brought many great benefit, therefore the defect rate of Thuan Phuong Group has decreased from 2.5% to 1.8. % (Figure 13).

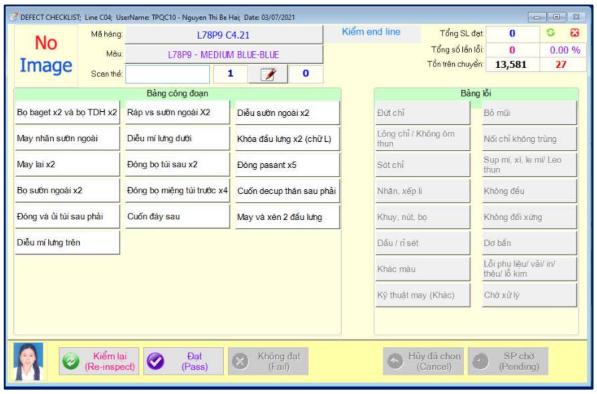


Figure 11: Step 2 in the operating procedure of Product Quality Management System

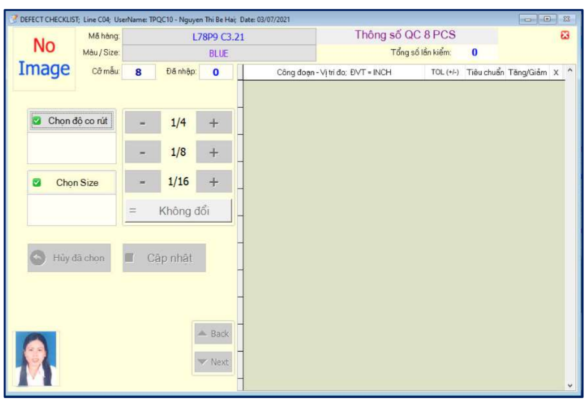


Figure 12: Step 3 in the operating procedure of Product Quality Management System

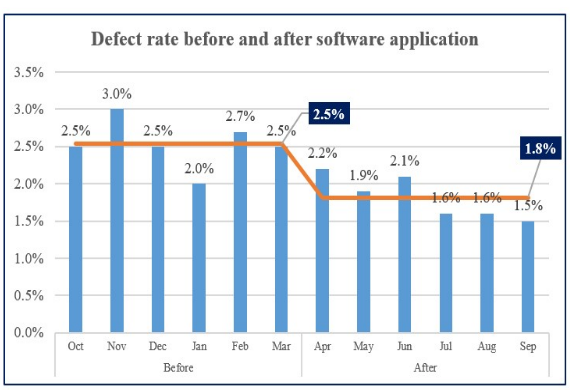


Figure 13. Defect rate before and after software application

6. Conclusions and Recommendations

6.1. Conclusions

In the current situation, improving product quality is one of the strategies to survive and develop enterprises in a sustainable way, as well as an important key for promoting economic integration and engagement, expanding international diplomatic of enterprises.

Product quality is decisive to improve compete ability, asserting the position of products, goods and brands of enterprises in the market. Thuan Phuong Group has comprehended it and go ahead to apply industry 4.0 to aim to automate its quality management process. Through this, building Raw Material Quality Management System software for fabric control to manage the input materials quality and Product Quality Management System software to manage the out put of product quality. After implementing the software in daily work, the defect rate decreasing from 2.5% to 1.8%. In addition to saving 20% QC department working time.

6.2. Recommendations

Thuan Phuong Group: Raw Material Quality Management System and Product Quality Management System have been implemented and brought many positive benefits to businesses. Therefore, company should consider investing in this system for some other process in the factory such as the process of checking materials, cutting fabric ... Vietnamese Company: Technology is advancing day by day, the application of automation based on 4.0 technology to production management in general and quality management in particular is a Turning Point that brings great success to Vietnamese enterprises. Therefore, Vietnamese enterprises should focus on investing in this field.

Reference

Edwards, C.D. (1968). *The Meaning of Quality.* Quality Progress, October, 37.

Gilmore, H.L. (1974). *Product Conformance Cost.* Quality Progress, Vol. **7**, Issue 6, 16 –19.

Broh, R.A. (1982). *Managing Quality for higher profits*. McGraw Hill, New York.

ASTM D5430 (2017), *Standard Test Methods for Visually Inspecting and Grading Fabrics*.

Cihat Okan Arıkan (2019). *Developing an Intelligent Automation and Reporting System for Fabric Inspection Machines*. TEKSTİL VE KONFEKSİYON, Vol: **29**, NO. 1.

Mominur Rahman (2015). *Study on Frequently Occurring Fabric Inspection Faults and Causes & Remedies*. Daffodil International University.

Khalid Ferdous (2020). *Study on Finished Fabric & Knit Garments Inspection Report*. Daffodil International University.

Kazi Rezwan Hossain (2019). *Study on Fabric & Garments Inspection of a woven Garments Factory*. Daffodil International University.

Suvir Kumar (2007). *A Computerized Greige Fabric Inspection System*. The Assumption University Library.

Abdullah Al Mamun (2018). *A Report on fabric inspection, faults, causes and remedies of frequently occurring 100% polyester fabric*. Daffodil International University.

Mazedul Islam (2013). *Minimization Of Reworks In Quality And Productivity Improvement In The Apparel Industry.* International Journal of Engineering and Applied Sciences, January 2013. Vol. **1**, No.4

Jung-Cheol Kim (2020). *A Study on Smart Factory Construction Method for Efficient Production Management in Sewing Industry.* J. lnf. Commun. Converg. Eng. 18(1): 61-68.

Engr.Mohammad Abdul Baset (2018). *Course Title: Study on Cutting, Sewing and Finishing Section at a Garments industry*. Daffodil International University.

Mohammad Faizur (2009). *Quality management in garment industry of Bangladesh.* Management of Sustainable Development. 2009, Vol. **1,** Issue 2, p1-7.

BIODEGRADABLE SUBSTANCES PREDICTION IN DRUG DISCOVERY USING RANDOM FOREST

Van T.H. Nguyen[1], Son V.T. Dao[2]

School of Industrial Engineering and Management, International University, Vietnam National University Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam

Email: [1]nthv.287@gmail.com, [2]dvtson@hcmiu.edu.vn

Abstract. Biodegradable substances have played an integral role in drug delivery applications due to several advantages in the pharmaceutical industry, and the quantitative structure-activity relationship (QSAR) models are encouragingly using to predict biodegradable compounds. However, with an exponential increment of the material index, it is complicated and time-consuming for pharmacists to process and generate insights sufficiently. Therefore, in this study, we propose a Machine Learning (ML) model to enhance the diagnosis process of readily biodegradable compounds, which performs on a QSAR dataset from the UCI machine learning repository. The framework is a hybrid model of a Random Forest classifier combined with a wrapper-based feature selection utilizing Sequential Selection Algorithms to find an optimal feature subset and boost the model's predictive performance. The proposed framework provides practical solutions and reduces bias to the majority group, yielding an accuracy of 99%, 98% in terms of MCC and recall. Compare RF_SMOTE-ENN to the two latest models: Balance Random Forest and a hybrid SVM model; the proposed framework outperformed two of those models in all measurements, thus demonstrating the model's advantage and ability to deal with biodegradable prediction.

Keywords: Biodegradation Prediction; Wrapper Feature Selection; Random Forest Classifier; Sequential Selection Algorithms; SMOTE-ENN

1. Introduction

A drug is an organic molecule that can produce a biological effect to inhibit disease. The process of researching and developing new drugs, from detecting drug-target interactions to putting them on the market, is challenging, costly, and time-consuming work which has a more than 50% probability of drug fails in clinical development due to low efficiency (Hwang et al., 2016). Therefore, controlling drug release that helps deliver drugs locally and closer to their target (drug efficacy proves) is crucial. The choice of using polymer, especially biodegradable polymers, has received much attention and become one of the most concerning issues in drug discovery since the polymeric system can protect the drug, cutting down on danger to the physiological environment, enhancing drug stability and bioavailability (Mousavian & Masoudi-Nejad, 2014).

Recently, the quantitative structure-activity relationship (QSAR) models, which can deduce chemical properties from their structure without experimentation, are encouragingly using to predict biodegradable compounds. However, due to the exponential increment of information in the dataset, many generated redundant instances make it complicated and time-consuming to process. Furthermore, many machine learning approaches support QSAR in determining biodegradable substances but experiment with all features causing probabilities of long-running time and biased results. As a result, this study aspires to minimize the extreme miss-classifications in the biodegradation prediction using a classification modelling method coupled with a feature selection algorithm to select the best subset features.

The structure of this paper is as follows. The literature review is introduced in Section II; next, sections Ⅲ and Ⅳ specify the methodology and experimental result, respectively. The conclusions are in the final section, which is section Ⅴ.

2. Literature Review

2.1. Application of machine learning in biodegradation prediction

QSAR's objective model is the meaning between molecular geometry and numerous mechanical, synthetic, and biological activities (Troopship, 2010). Recently, many researchers have taken part in developing a machine learning model to support QSAR models in biodegradability prediction. Tang et al. (2020) applied multiple linear regression (MLR) and support vector regression (SVR) with inputs including 171 molecular descriptors evaluated by the Root Mean Squared Error (RMSE), which found that SVM achieved better performance with the RMSE of 0.08. Based on MITI and BIOWIN datasets, Cheng et al. (2012) had built SVM, K-nearest Neighbor (KNN), and C4.5 decision tree models combined four different feature selection techniques, gained the AUCs of 0.856, 0.844, and 0.873, respectively (Cheng, 2012). Mansouri et al. (2013) also analysed 1725 chemicals and proposed KNN and SVM models using 12 and 14 molecular descriptors demonstrating good performances, with all accuracies higher than 82% in training, test, and external validation sets (Mansouri et al., 2013). However, the developed models used less transparent modelling techniques (SVM and KNN) due to the black-box approach. Using the QSAR biodegradation dataset, Houssein, Essam H., et al. (2020) also implemented a hybrid model of SVM with CHHO–CS-Piece, a combination of Harris's Hawks Optimizer (HHO) with cuckoo search (CS) and chaotic maps, offering an accuracy of 0.84 (Houssein et al., 2020). As a result, SVM is one of the most widely used and powerful approaches in pharmaceutical applications. However, SVMs are trying to minimize error rates, resulting in a clear bias towards the majority class and poor performance in imbalanced datasets. To avoid the limitation of SVM, George & John (2020) performed data oversampled with Synthetic Minority Oversampling Technique (SMOTE) combined with feature selection and SVM to predicting biodegradable compounds of the SMILES dataset, achieving 0.941 of specificity (George & John, 2020). Another way to reduce bias in DTI is to switch over to tree-based algorithms as it tends to incorporate class weight and penalize the minority misclassification.

Chen et al. (2014) proposed and compared some of the biodegradability prediction models, which include Decision Trees (DTs), Functional Tree (FT), and Logistics Regression (LR) evaluated by overall predictive accuracy. They found that FT gave the best predictivity with accuracies of 81.5% and 81%, respectively, on the training set and test set. Elsayad et al. (2020) also implemented tree-based algorithms in the QSAR model to classify substance biodegradation and the ability to deal with unbalanced data. The comparison between Balance Random Forest (BRF), DTs, KNN, and SVM machine learning algorithms showed that BRF outperformed the typical classifications with AUROC of 0.92 (Elsayad et al., 2020) Investigating the chemical compounds dataset in a different area, Banerjee et al. (2018) also built a binary classification model to predict chemical compounds' bitter and sweet taste. They concluded Random Forest Algorithm achieved an accuracy of 96% and an AUC of 0.98 for bitter and sweet taste prediction in the independent test set (Banerjee & Preissner, 2018).

2.2. Feature selection in biodegradation prediction

Furthermore, having found many structural features in biodegradable compounds such as halogens, heavy atoms, esters, hydroxyl groups, Etc., only some can enhance biodegradability, causing the need to extract a subset of features to minimize the error rate and effects of noise data. Elsayad et al. (2020) implemented ranking techniques for feature selection by ordering their importance based on the p-value of F-statistics and likelihood ratio chi-squared, which returns the top 10 descriptors according to their contribution are also ones that have 100% importance by p-value [10]. However, since the biased result appears quite often in this technique, the returned subset might not be optimal and obtain redundant features. On the other hand, Yong Wang et al. proposed a backward elimination reply on the Random Forest algorithm in a binary classification of breast cancer resulting in the top 5 significant features obtained an accuracy of 100% and 99% on average (Nguyen et al., 2013).

However, most related studies disregarded the class imbalance problem, especially those that used accuracy instead of other evaluation metrics. Consequently, the study will apply a Random Forest Model using Wrapper Methods to predict the biodegradability of chemical descriptors, focusing on solving two critical questions in this dataset regarding feature selection and bias evaluation. Note that the impact of the re-sampling method is also explored.

3. Methodology

3.1. Data Description

In this study, a QSAR dataset of 1055 substances dataset provided by the Department of Earth and Environmental Sciences – the University of Milano-Bicocca on UCI machine learning repository, which included 41 chemical compound activities (Table 1), was used to discover the relationships between chemical structure and biodegradation of molecules. The detailed description of each feature and the dataset is available at: https://archive.ics.uci.edu/.

Table 1. Abbr. of features used in the dataset

Feature	Type	Feature	Type
nHM	Ordinal	nArCOOR	Ordinal
F01NN	Ordinal	nX	Ordinal
F04CN	Ordinal	Class	Ordinal
NssssC	Ordinal	SpMax_L	Continuous
nCb	Ordinal	J_Dze	Continuous
nCp	Ordinal	C	Continuous
nO	Ordinal	SdssC	Continuous
F03CN	Ordinal	HyWi_Bm	Continuous
nNN	Ordinal	LOC	Continuous
nArNO2	Ordinal	SM6_L	Continuous
nCRX3	Ordinal	F03CO	Continuous
nCIR	Ordinal	Me	Continuous
B01CBr	Ordinal	Mi	Continuous
B03CCl	Ordinal	SpPosA_Bp	Continuous
N073	Ordinal	SpMax_A	Continuous
B04CBr	Ordinal	Psi_i_1d	Continuous
nCrt	Ordinal	SdO	Continuous
C026	Ordinal	TI2_L	Continuous
F02CN	Ordinal	SpMax_Bm	Continuous
nHDon	Ordinal	Psi_i_A	Continuous
nN	Ordinal	SM6_Bm	Continuous

3.2. Feature Selection: Wrapper Method

Wrapper methods are applied to train a model and chose the subset giving the highest score of evaluation. The objective function of Wrapper Methods:

$$J(X) = \max_{J\subseteq y,\ |Z|=d} J(Z) \quad (1)$$

Where Y is the original set of features and X is the selected subset in d-dimensional feature space.

This paper mainly focuses on two Sequential Selection Algorithms, one of the most common uses in Wrapper: Sequential Forward Selection and Sequential Backward Selection. The Sequential Feature Selection (SFS) algorithm, from a null set, will seek the feature giving the highest score of the objective function and add it to the set. From the second step, SFS is adding a new feature independently to expand the current subset then evaluate the performance of the new one. In contrast, Sequential Backward Selection (SBS) algorithm goes from a complete set of variables and excludes the minor contribution feature, the worst one, in predictor performance. These processes repeat until meeting the required number of features (Nguyen, 2013).

3.3. Classification Algorithm: Random Forest

Random Forest Random Forest (RF), an ensemble learning method, comprises multiple decision trees to train data and make classification or regression predictions by the trees' majority vote (Fig.1).

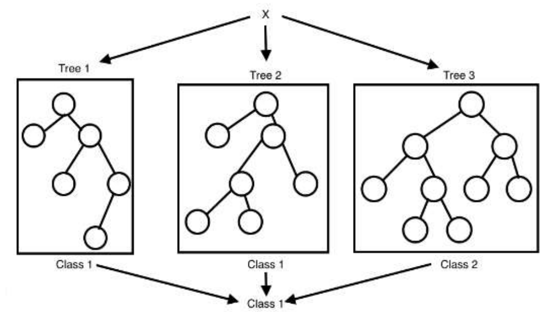


Figure 1. A framework of Random Forest Classification

This technique is much more robust and accurate than a single Decision Tree, which reduces bias and overfitting results. Firstly, training every decision tree on a different subset of samples, then at each tree node, the tree will search through the features and choose ones resulting in the most significant reduction Entropy to split out the nodes. Entropy is defined as

$$Entropy(S) = -\sum_{i=i}^{C} P_i(logP_i) \quad (2)$$

Where S: the subset of training samples, P_i : the percentage of negative and positive samples in S
Then, after several splits, the final prediction of each tree is made at the leaf node, and the decision of Random Forest will base on the most prediction solutions. In particular, Random Forest uses the formula:

$$R(x) = \frac{1}{T}\sum D_t(x) \quad (3)$$

where $D_t(.)$ is a decision tree, and T is the number of trees in the Random Forest. Indeed, $D_t(x)$ is the score that a single tree t predicts for a vector x.
The objective function of Random Forest is:

$$maximize_{x,q_{i,j}^t} R(x) \coloneqq \frac{1}{T}\sum D_t(x) \quad (5)$$

where t= 1…T with T is the number of trees in the forest
$i \in N^t$ with N^t is the set of interior nodes in tree t

$j \in L^t$ with L^t is the set of leaves in tree t
$q^t_{i,j}$ is a binary variable such that $q^t_{i,j} = 1$ if and only if the solution x is in a leaf j that is a descendant of node i

3.4. SMOTE-ENN

An imbalanced class may be challenging for ML algorithms by causing a risk of bias toward the majority class. Hence, re-sampling techniques, such as oversampling and under sampling, are well known to equal the proportion of both majority and minority classes. Note that SMOTE, which is an oversampling method, randomly picks a minority substance and generates synthetic samples by gathering information about this selected and its nearest neighbour.
In this case, we will only focus on SMOTE ENN (SMOTE with Edited Nearest Neighbour) developed by Batista et al. (2004), which allows removing observations that do not belong to any class.

3.5. Proposed Method

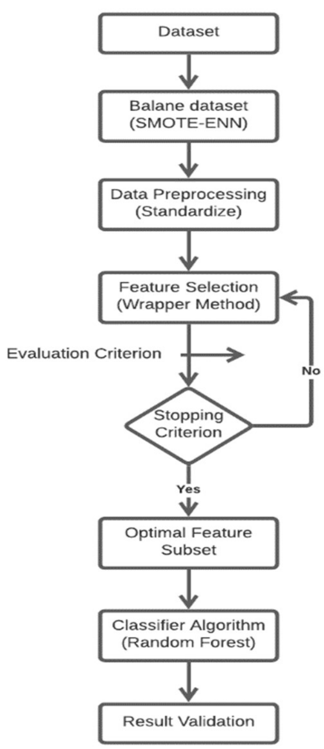


Figure 2. The flow chart of the upgrade framework

The proposed method in Fig.2 is as follows:

Step 1: Implement SMOTE-ENN to balance the proportion between negative and positive classes.

Over 60% of the target variable non-biodegradable (NRB) (699/1055 compounds) showing a slight class imbalance. In a different circumstance, non-biodegradable substances account for over 70% (479/670 compounds) in the external dataset, which is approximately 5% higher and more imbalanced than the predictive one.
At first, SMOTE will randomly select an instance of the readily biodegradable (RB) class (x_i) and estimate the difference between it and its K-Nearest Neighbour (K_{xi}) using Euclidean distance, then multiply by a random number in a range of [0,1] and result in a synthetic data sample (x new) by the following formula:

$$x_{new} = \left(\hat{x_i} - x_i\right) \times \delta \quad (6)$$

Where $\hat{x_i}$ *is an element in* K_{xi}; $\hat{x_i} \epsilon X_{min}$

The formula to calculate Euclidean distance is:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (7)$$

When the proportion of two classes is balanced, SMOTE-ENN will generate the neighbourhood cleaning rule based on the edited nearest-neighbour (ENN) to remove samples that differ from its K-nearest neighbours' class. Note that a default K=3 will be applied for calculation.
The dataset is now more balanced than the previous version, including 55.59% of RB compounds and 44.41% of NRB ones.

Step 2: Standard Scaler to centre data around 0 and 1 (Pre-processing data)

As there is no missing value in the dataset, the balanced set then will be standardized using the below formula:

$$z = \frac{x - \mu}{\sigma} \quad (8)$$

where μ: mean and σ: standard deviation.

Moreover, because the zero values are common among these columns, special "functional" groups in the molecule affect their properties and chemical behavior. Besides, when zero is a valid value for a variable, other extreme values can easily consider as outliers. So, in this case, we will not consider outliers yet and figure out whether it has any valuable contribution.

Step 3: Feature Selection using SFS and SBS method

Step 3.1: Filter out irrelevant variables.

Firstly, applying the Correlation Matrix to drop high correlated continuous variables. Noted that two different features are strongly related if their correlation is higher than or equal to 0.7.

$$R(i) = \frac{cov(X_i,Y)}{\sqrt{var(x_i)*var(Y)}} \quad (9)$$

Where x_i: the i^{th} variable, Y: predicted output.
Based on the computational results, nine pairs of high correlation are identified. However, based on the distribution of the features and how they correlated with each other, we only eliminate TI2_L, SpMax_B (m), J_Dz(e), and Me.
Then applied Chi-square for categorical ones to estimate the ability to specify classes. Within a 95% confident interval, any p-value greater than 0.05 proves that the variable is independent of the target.

H_0: There is no relationship exists between feature and target.
H_1: There is no relationship exists between feature and target.

After conducting the statistic test, only nN-N, which shows a p-value of 0.128, is rejected from the dataset. Therefore, there are 37 attributes left in the classification set.

Step 3.2: Create new features.

Based on chemical background knowledge, we decide to create new features to support dividing the target classes. Specifically, the number of heavy atoms in a compound (nHM), with none, one, or more than one heavy atom, is specified as "light", "functional", and "heavy", respectively. The number of terminal primary C(sp3) (nCp) at 0, 1, 2 or greater than two are defined as "ring", "semiring", "linear", and "branched", respectively.

Step 3.3: Feature selection

Finally, apply SFS and SBS methods for feature selection to select an optimal subset of attributes. In this study, we use *SequentialFeatureSelector()*, which is available in mlxtend library, an online package containing most wrapper methods techniques to implement these two algorithms. We also set up a Standard Random Forest as an ML-based algorithm for the feature selection process. In this study, we use F1-score to evaluate the model performance on the test set; thus, SBS and SFS try to select features that return a maximum F1-score. (The define and formula of the F1-score will be discussed in Step 5).

Step 4: Classification model using standard Random Forest (random state = 42) and the selected subset

Data sets are randomly split into 80% training set and 20% test set, ensuring the initial distribution of the distinct target class (RB=1 and NRB=0). Stratified 5-fold cross-validation is applied to ensure each fold acts as a test set at some point and avoids biased results during the training process.

Step 5: Result validation

In this paper, F1-score, and Matthews Correlation Coefficient (MCC) will be chosen to evaluate the performance of our proposed method.
Recall determines how well the model predicts each class, while precision answers how confident the model output is, which is calculated by these formulas:

itive NP: Negative Positive FN: False Negative TN: True Negative

F1-score is a metric that depends on the value of precision and recall and is defined as follows:

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recal} \quad (12)$$

In binary classification, another evaluation technique, called the phi-coefficient (φ) or Matthews Correlation Coefficient (MCC), treats it as a multiclass problem and calculates its correlation coefficient. The stronger correlation between the predicted variable and the valid positive variable is, the better the model will be. Computing the MCC is:

$$MCC = \frac{TP*TN-FN*FP}{\sqrt{(TP+FN)*(TP+F\ \ )*(TN+F\ \ )*(TP+FP)}} \quad (13)$$

Besides, we also use specificity and sensitivity to compare with other published methods. These measurement formulas are:

$$Specificity = \frac{TN}{TN+F} \quad (14)$$

$$Sensitivity = Recall \quad (15)$$

4. Results and discussion

4.1. Performance of the model

In Table 2, the SFS subset has both outperformed the SBS subset before and after re-sampling, giving the most accurate prediction in terms of F1-score and accuracy, respectively. Moreover, it verifies that the re-sampling dataset by SMOTE-ENN enhances the classification performance for all methods and progresses by leaps and bounds of all now measurements. The MCC has increased from 0.69 to 0.98, up to 29%, indicating a stronger correlation between prediction and observation and better performance.

Table 2. Summary of model performance

Data	Model	Number of features	F1-score	MCC	Accuracy
Main dataset	Baseline model	42	0.78	0.69	0.86
	Using SFS subset	**24**	**0.77**	**0.69**	**0.86**
	Using SBS subset	14	0.76	0.66	0.85
External test set	SFS subset	24	0.74	0.65	0.86
Main dataset + SMOTE-ENN	Baseline model	42	0.98	0.96	0.98
	After adding new features and remove irrelevant features	42	0.98	0.97	0.98
	Using SFS subset	**24**	**0.99**	**0.98**	**0.99**
	Using SBS subset	24	0.98	0.97	0.98
	Using SFFS subset	24	0.98	0.97	0.98
External test set + SMOTE-ENN	SFS subset	24	0.95	0.9	0.95

Besides, we are trying to minimize false-negative (FN) in the pharmaceutical industry as it leads to more risks than false-positive (FP). For example, when an active compound is declared defective, it just causes a business loss by rejecting a material that should be used. Otherwise, if a defective material is told faultless, it is a false negative and can harm the patient when launching the product. In the RF_SMOTE-ENN model, there is no FN and only two FP (Fig. 3), indicating only a 0% false-negative rate and 2.25% false-positive rate.

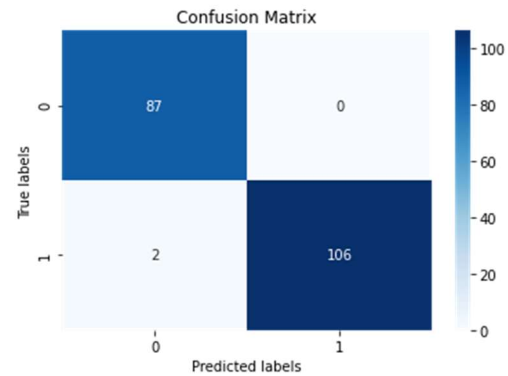


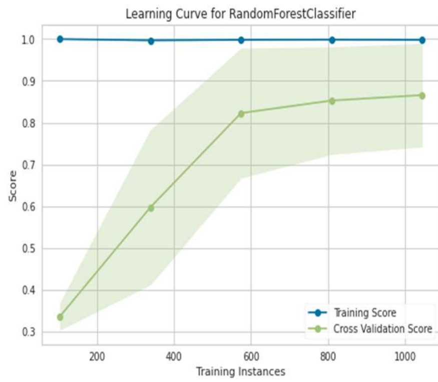
Figure 3. The confusion matrix using the subset of SFS+SMOTE-ENN

When driving deeper into each category (Table 3), the model achieves 29% more recall of the positives, which successfully predicts 98% of the minority class and gives 100% precision. Likewise, the precision score of the negatives has increased from 0.86 to 0.98, up 12%, and by 4% in terms of recall. Thus, the model with SMOTE-ENN has marked growth in all metric scores compared to the initial Random Forest algorithm, showing the benefits of this re-sampling method.

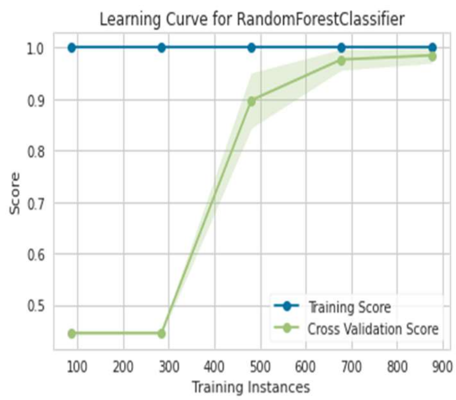
Table 3. The confusion matrix by class categorical

Is biodegradable	Subset	Precision	Recall	F1 score
1	**SFS**	0.86	0.69	0.78
	SFS_SMOTE-ENN	1	0.98	0.99
0	**SFS**	0.86	0.96	0.91
	SFS_SMOTE-ENN	0.98	1	0.99

Moreover, Fig.4 and Fig.5 have proved a complete dominance of the RF_SMOTE-ENN model over Random Forest single models. The accuracy is up by approximately 13% accuracy (from 0.86 to 0.99) and 16.7% in terms of AUC (from 0.823 to 0.997). Note that the AUC stands for distinguishing between biodegradable compounds and non-biodegradation ones. The higher AUC, the better the predictive performance of the model.



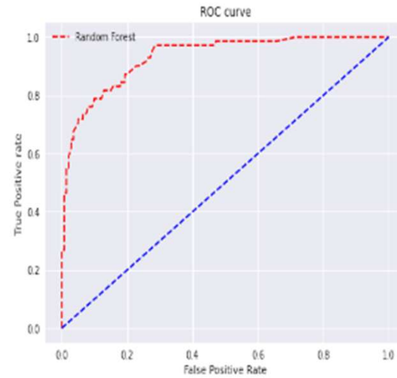
(a) RF_SFS



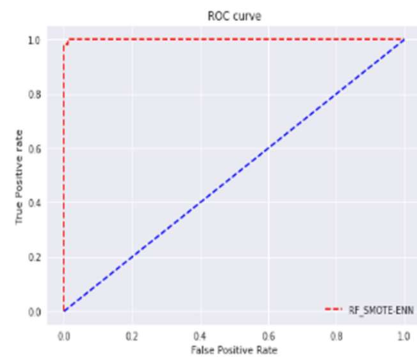
(b) RF_SMOTE-ENN_SFS

Figure 4. Accuracy performance of the two best models

Fig.5 shows the classifier performances of two algorithms, in which a suitable model has fewer false positives values and is more sensitive will return a higher point on the upper left corner of the ROC chart.



(a)



(b)

Figure 5. Receiver Operating Characteristic (ROC) (a). FR_SFS and (b). RF_SMOTE-ENN_SFS

4.2. Feature selection

Table 4 will show the top 5 feature importance and how each feature contributes to the prediction of the two subsets coming up with the most accurate classification results before and after applying SMOTE-ENN. The features importance method is available on the sklearn package of implementation of the Random Forest as permutation_importance, which randomly shuffles each feature and assigns feature that impacts the most on the performance as the important one.

According to Table 4, it is interesting to note that all these top 5 descriptors have a good separation between the two classes of the target, confirming their importance when predicting biodegradable compounds. SpMax_L and SdO are in the two models (RF_SFS and RF_SMOTE-ENN_SFS), which verify their essential roles in classifying the biodegradability of substances. Specifically, SdO relates to an oxygen functional group that helps boost biodegradability, while SpMax_L, which is collected from the Laplacian matrix, represents the molecular branching, and relates more to the NRB molecules.

Table 4. Top 5 essential features selected by two methods.

Rank	RF_SFS	RF_SMOTE-ENN_SFS
1	SpMax_L	SM6_B(m)
2	nHM	F02[C-N]
3	nO	nHM
4	SdO	SdO
5	nN	SpMax_L

Moreover, the matrix-based descriptor like SM6_B(m), molecule informing on the number of substituted benzenes F02[C-N], nN in nitrogen functional group always related to NRB compounds, while nO is related more to RB substances (Mansouri et al., **2013). Note that we overlook nHM because its role had alread**y been mentioned before. On the other hand, new features such as mS_branched and nHM_functional are also chosen as essential features, indicating their contribution to the prediction model. Table 5 shows all features in the two subsets of RF_SFS and RF_SMOTE-ENN_SFS.

Table 5. Selected feature subsets

Method	Feature Subset
RF_SFS	*B01[C-Br], B03[C-Cl], C%, F01[N-N], F03[C-N], F03[C-O], F04[C-N], HyWi_B(m), N-073, nArCOOR, nArNO2, nCb-, nCIR, nCRX3, nHDon, nHM, nN, nO, NssssC, nX, Psi_i_1d, Psi_i_A, SpMax_L, nHM_functional*
RF_SMOTE-ENN_SFS	*B01[C-Br], B04[C-Br], F01[N-N], F02[C-N], F04[C-N], HyWi_B(m), LOC, nArCOOR, nArNO2, nCb-, nCp, nCrt, nHDon, nHM, nN, NssssC, Psi_i_1d, Psi_i_A, SdO, SdssC, SM6_B(m), SpMax_A, SpMax_L, mS_branched*

4.3. Model Comparisons

According to Table 6, RF_SMOTE-ENN gives the highest score, Balance Random Forest (Elsayad et al., 2020) that bootstraps classes in the same proportion, is the second rank, and a hybrid model of Support Vector Machine with CHHO–CS-Piece (Houssein et al., 2020) is at the bottom. The result is relatively consistent with the previous claim that algorithms focusing on minimizing the error rate like SVM may ignore the imbalance and result in bias, while tree-based algorithms that combine class weight can deal better with imbalanced sets and reduce misclassification of minority groups. This claim also fits with the recall value of SVM_CHHO–CS-Piece, only 52.5%, showing that the model has failed to predict nearly half of the minority samples.

Compare RF_SMOTE-ENN to the top-performing published models using Balance Random Forest and a hybrid SVM model; RF_SMOTE-ENN with SFS subset outperforms two of those models in all measurements. As expected, our method achieves 12% more accuracy than Elsayad et al. (2020) and 11% in terms of ROC. Moreover, when compared to Houssein et al. (2020), RF_SMOTE-ENN accomplishes approximately as twice as sensitivity and 28.4% more specificity than SVM_CHHO–CS-Piece. Thus, RF_SMOTE-ENN gives the best results among the three methods.

Table 6. Statistical results of individual classification models on the QSAR dataset

	RF_ SMOTE-ENN	Balanced RTs	SVM_ CHHO–CS-Piece
F1 score	0.99	0.82	0.66
MCC	0.98	0.73	x
Accuracy	0.99	0.87	0.84
Recall/Sensitivity	0.98	0.92	0.525
Precision	1	0.75	0.72
Specificity	0.977	0.95	0.693
ROC	0.99	0.88	x

4.4. Social, economic, and environmental impacts analysis

Drug discovery is a challenging, costly, time-consuming, and easily failed process. According to the report "Measuring the return of pharmaceutical innovation" of Deloitte in 2018, the R&D process costs were double in the past eight years, while the average sales dropped by a half. The reason is the long processing time and the high failure rate of a drug candidate in terms of specificity and efficacy.

However, Machine Learning (ML) has offered biopharma companies a more accurate, speedy drug discovery. ML solutions help reduce the pharmaceutical industry productivity crisis by supporting scientists to test diverse hypotheses faster with a lower cost, result in both scientific and economic benefits. ML also assists in identifying toxicity reactions, compound selection, and target prediction, which promises an unbiased insight into drug efficiency and safety and decreases the delivery time of drugs to patients. Moreover, ML can replace unnecessary and causing pain, suffering, and death experiments on the animal, stop wasting on animal life, affection of the ecosystem, and protect the environment.

Furthermore, with the outbreak of COVID-19, ML-driven drug discovery takes advantage of its speed and accuracy to identify and propose coronavirus treatment, which represents a savior in the war against the pandemic (Artificial Intelligence in Drug Discovery Market, 2020-2027) (Vatsal, 2021). Thus, investment in biotechnology dramatically rises in the pandemic, promising new opportunities for startups and help produce more drugs for patients.

5. Conclusion

This paper conducted a hybrid model that combines a Random Forest classifier with the Wrapper feature selection technique and re-sampling method SMOTE-ENN to classify degradable materials without any chemical experiments. The proposed framework provides practical solutions and reduces bias to the majority group, yielding an accuracy of 99%, 98% in terms of MCC and recall. Compare RF_SMOTE-ENN to the two latest models: Balance Random Forest and a hybrid SVM model; the proposed framework outperformed two of those models in all measurements, thus demonstrating the model's advantage and ability to deal with biodegradable prediction. Besides, feature selection is used to get the frequent occurring features in RB ad NRB compounds, reducing the computation demand while maintaining accuracy outputs, of which the SpMax_L and SdO are two of the most important features affecting prediction output. From this study, it could be inferred that NRB substances, as well as accessible toxicity compounds, do exhibit specific features regarding heavy atoms, nitrogen functional group, ... This model will be essential in drug discovery to define and eliminate those materials causing harm to patients, improve drug efficacy, and decrease the failure rate in clinical trials.

However, the present study has only investigated the QSAR model dataset. In future work, we will apply more molecules libraries to find other new candidates and patterns in biodegradable substances and enhance model performance. This framework is the potential application to real-life drug discovery in Stage 1 and becomes a supporting tool for scientists and pharmacists. It can also predict degradable materials using in industry, helping to reduce the amount of plastic waste entering the environment and producing less harmful products with a shorter degradation time.

Reference

Banerjee, P., & Preissner, R. (2018). *Bittersweet Forest: a random forest-based binary classifier to predict bitterness and sweetness of chemical compounds.* Frontiers in Chemistry, 6, 93.

Chandrashekar, G., & Sahin, F. (2014*). A survey on feature selection methods.* Computers & Electrical Engineering, **40**(1), 16-28.

Chen, G., Li, X., Chen, J., Zhang, Y. N., & Peijnenburg, W. J. (2014). *Comparative study of biodegradability prediction of chemicals using decision trees, functional trees, and logistic regression.* Environmental toxicology and chemistry, **33**(12), 2688-2693.

Cheng, F., Ikenaga, Y., Zhou, Y., Yu, Y., Li, W., Shen, J., ... & Tang, Y. (2012). *In silico assessment of chemical biodegradability.* Journal of chemical information and modeling, **52**(3), 655-669.

Elsayad, A. M., Nassef, A. M., Al-Dhaifallah, M., & Elsayad, K. A. (2020). *Classification of Biodegradable Substances Using Balanced Random Trees and Boosted C5. 0 Decision Trees.* International journal of environmental research and public health, **17**(24), 9322.

George, A., & John, M. (2020, October*). Impact of Oversampling on the Classification of Readily Biodegradable Materials.* In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) (pp. 1-5). IEEE.

Houssein, E. H., Hosney, M. E., Elhoseny, M., Oliva, D., Mohamed, W. M., & Hassaballah, M. (2020). *Hybrid Harris's Hawks Optimization with Cuckoo Search for drug design and discovery in cheminformatics.* Scientific Reports, **10**(1), 1-22.

Hwang, T.J., Carpenter, D., Lauffenburger, J.C., Wang, B., Franklin, J.M. and Kessel Heim, A.S., 2016. *Failure of investigational drugs in late-stage clinical development and publication of trial results.* JAMA internal medicine, **176**(12), pp.1826-1833.

Mansouri, K., Ringsted, T., Ballabio, D., Todeschini, R., & Consonni, V. (2013). *Quantitative structure-activity relationship models for ready biodegradability of chemicals.* Journal of chemical information and modeling, **53**(4), 867-878.

Mousavian, Z., & Masoudi-Nejad, A. (2014). *Drug–target interaction prediction via chemogenomic space: learning-based methods.* Expert opinion on drug metabolism & toxicology, **10**(9), 1273-1287.

Nguyen, C., Wang, Y., & Nguyen, H. N. (2013). *Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic.*

Reports and Data, H. (2020). *Artificial Intelligence in Drug Discovery Market, 2020-2027*. Retrieved from https://www.reportsanddata.com/report-detail/artificial-intelligence-in-drug-discovery-market

Tang, W., Li, Y., Yu, Y., Wang, Z., Xu, T., Chen, J., ... & Li, X. (2020). *Development of models predicting biodegradation rate rating with multiple linear regression and support vector machine algorithms.* Chemosphere, 253, 126666.

Troopship, A. (2010). *Best practices for QSAR model development, validation, and exploitation.* Molecular informatics, **29**(6-7), 476-488.

THE STUDY AND DESIGN OF CRM SYSTEM IN ADMISSIONS COUNSELING AND SUPPORT FOR STUDENTS: CASE OF INTERNATIONAL COLLEGE OF HO CHI MINH CITY

Nguyen Quynh Lam[1], Do Doan Trang[2], Vo Trong Cang[3]

Tang Minh Huong[4], Nguyen Thi My Le[5]

[1]*Binh Duong University, Binh Duong, Viet Nam, nqlam@bdu.edu.vn*

[2]*Binh Duong University, Binh Duong, Viet Nam, doantrang.bolt@bdu.edu.vn*

[3]*Binh Duong University, Binh Duong, Viet Nam, vtcang@bdu.edu.vn*

[4]*Dai Viet My English Center, Dak Nong, Viet Nam, minhhuongtang@gmail.com*

[5]*International College Of Ho Chi Minh City, Ho Chi Minh, Viet Nam, lentm@ich.edu.vn*

Abstract. Nowadays, admissions counseling and support for students are the most important elements at college. Parents and students have been considered as clients are experiencing training service at college, so the care and support for parents and students are a necessity. The study and design of CRM system has introduced more efficient methods in case of communication with clients and support admissions counselor department to improve productivity. On the other hand, CRM system helps the interaction between admissions counselor department and others in order to reduce students drop out in the middle of course. In addition to CRM system also help students have experienced free course and have got information before entering the course at college. The conclusion for application the CRM system in admissions counseling and support for students has increased the number of incoming students by 35% in (2018 -2019) and reduced the dropout rate by 10% in the same period last year.

Keywords: CRM; Customer Relationship Management; admissions counseling; support for students

1. Introduction

CRM has been described as "a customer-focused business strategy that aims to increase customer satisfaction and customer loyalty by offering a more responsive and customized service to each customer" (Croteau, A. and Li, P, 2003). Students, students' parents, short-term students, businesses, school staff (employees and lecturers), and school partners are all examples of "customers" in the education sector. Thus, the customer relationship in colleges and vocational schools includes many other relationships besides student and school, such as school - parent, school - lectures, school - high school teachers, schools employees, employees-students, lecturers-students, and schools-enterprises. However, the primary relationship that the CMR system in education must develop and constantly update is. Because the system collects information at all stages of interaction with customers (parent/potential student/main student), the educational CRM system assists the school in having a comprehensive understanding of the "customer" needs (Abubakar, M., Mukhtar, S, 2015). As a result, good relationship management will provide the following benefits:

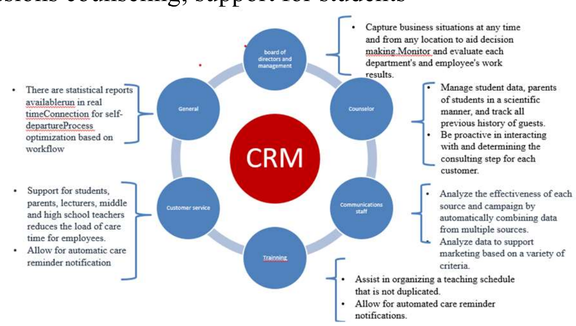


Figure 1. Benefits of CRM

The focus of marketing today is not so much on attracting "buyers" as it is on customer satisfaction, getting them what they want, and cultivating a relationship with them that is more than just commercial. As a result, the CRM system of educational institutions is an appropriate tool for the school to grasp the needs of students in terms of training, allowing the school to offer courses and services that are appropriate to the students' characteristics and requirements, allowing the school to improve and achieve better outcomes to reduce dropouts (Rigo, 2016) (Adikaram, 2016) (Abubakar, M., Mukhtar, S, 2015) (Wahab, 2011).

Because CRM systems are industry-specific (Stair, 2010), organizations from various industries will develop distinct CRM models. In the same field, but the CRM system is not used equally, so developing a CRM system that is aligned with business strategies is a prerequisite for success. As a result, in order to implement a CRM system in education in accordance with the enrollment counseling and student support process at Vietnam's colleges and intermediate schools, a thorough understanding of the current enrollment and survey process is required. status of colleges and intermediate schools, from which the appropriate CRM system for the school's strategy can be applied. Understanding the realities of admissions counseling and student support at colleges and intermediate schools, my partner and I designed a CRM system for use in admissions and support departments. Help students at International College Of Ho Chi Minh City (ICH) to achieve transformative results.

Based on the analysis of the preceding studies and information, the following structure will be used to present the research contents and design of a CRM system in enrollment counseling and student support: Part 2: Objectives; Part 3: Research methods; Part 4: CRM research and education application; Part 5: Current status of CRM application in International College Of Ho Chi Minh City (ICH); Part 6: At International College Of Ho Chi Minh City, the CRM process is proposed for improvement; Part 7: CRM solutions and achievements contents; Part 8: Conclusion.

2. Objectives

Conduct research on the concept of CRM, the state of CRM application around the world, and the current state of customer relationship management at the ICH's admissions and customer support departments.

Evaluate and apply the necessary and appropriate business processes to the International College Of Ho Chi Minh City (ICH).

Create a CRM system that is compatible with social networking applications.

Expanding deployment and operation at private colleges and intermediate schools.

3. Research methods

Qualitative research method: The research collects and selects published documents, articles in scientific journals, and official documents on reputable websites related to the research issue, as well as surveys the current status of CRM application at some universities around the world and some educational institutions in Vietnam.

System development method: Using the system development method (SDM - System Development Method (Nunamake, 1991) and the findings of qualitative research, the study will continue to conduct surveys about the current status and needs at the International College in order to determine the requirements for the CRM system.

4. CRM research and education application

Overseas: According to research conducted by the AACRAO organization on US universities (AACRAO, 2014). According to the findings of this report, 64% of universities have implemented CRM. The remaining 42 percent are not currently using CRM but are considering it. Another study conducted in Indonesia on the relationship between loyalty, satisfaction, and CRM systems (Zani, 2013) discovered that good customer relationship management increases satisfaction. Not only that, but student satisfaction makes students feel proud, piques their interest, and encourages them to study.

In Vietnam: The CRM system has been widely used in industries such as tourism, telecommunications, and sales, and CRM has demonstrated its role in "Finding, Closing, and Keeping" customers. CRM, on the other hand, is not widely used in the field of education, and is primarily used by English centers, international schools, and overseas study companies such as Apollo, Wellspring and Instulink.The system can promote customer closure and manage and evaluate the working capacity of consultants for Apolo English Center. The system assists Wellspring School in better managing students, connecting functional departments, synchronizing data, and saving time. In short, CRM in education in Vietnam is underutilized and has few features.

5. Current status of CRM application in International College Of Ho Chi Minh City (ICH)

According to Abdullah's research scale on the HEDPERF scale (Abdullah, 2006), it includes the necessary factors to help students be satisfied, including factors related to employees' customer service duties. (by non-academic staff) such as: staff service capacity, how to solve departmental staff problems, and dedicated advice to students. This scale is also used to test the reality of student satisfaction at the International College of Ho Chi Minh City (ICH) (Lam, 2015). This scale is also used at the International College of Ho Chi Minh City (ICH); to assess the reality of student satisfaction. According to the findings of a 500-student sample study, the factor "Student support services" (the factor about student care) is an important factor influencing the satisfaction of International College students (Figure 2)

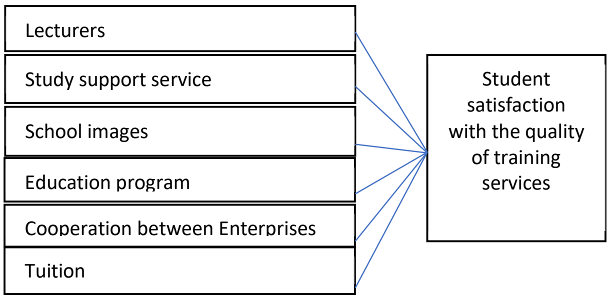


Figure 2. Model for assessing the quality of training services provided by ICH

With this level of interest, we can see that developing a CRM system to manage customer relationships is critical. As a result, the International College has used CRM in student management since the 2016-2017 school year, and it includes the following CRM components:

- *Student management:* This is the profession of the training department and the student affairs department that is used to manage grades, attendance, and personal information.
- *Management of lecturers and staff:* The administrative and human resources department is used to manage information about the lecturer's profile, employee information, schedule scheduling, and teaching schedule.

With such a simple CRM system as described above, only managing basic problems rather than using Excel software did not promote the CRM system's effectiveness. CRM does not include marketing activities, enrolment consulting, customer care (potential, official), or general reports. As a result, the author proposes a number of constraints that must be overcome, specifically as follows:

- *Marketing activities:* No specific marketing strategy has been demonstrated, no potential and official data sources of potential students have been gathered and analysed, no students have been classified by industry, and no potential student needs have been identified.
- *Recruitment consulting activities:* Ineffective data management, inability to track customer history, inability to synchronize data, resulting in duplicate or incorrect data, inability to control the frequency of potential customer care (forgetting customers or unintentionally disturbing customers).
- *Supportive care activities:* There is no process in place to care for "customers" during the school day; the care is uneven, time-consuming, and ineffective, and it cannot result in customer satisfaction.
- *Reporting activities:* Because there is no timely and detailed reporting mechanism, management is slow to grasp the business situation, cannot monitor and evaluate employee work results, and cannot evaluate employee effectiveness. strategies for marketing.

Because the benefits of the CRM system used in the previous year (2016-2017) did not meet the needs of the College of Economics, the College of Economics implemented an upgraded CRM system in the current academic year (2018-2019) and achieved positive results.

6. Current status of CRM application in International College Of Ho Chi Minh City (ICH)

CRM systems are unique to each industry, and even within the same industry, system characteristics will differ. Employee management program, employee recruitment and retention system, student care system, program marketing and enrollment, campaign management program, lead tracking and consulting process, student information system, and social network management system are among the ten supporting functions provided by EliNext's CRM system solution (Elinext, 2021).

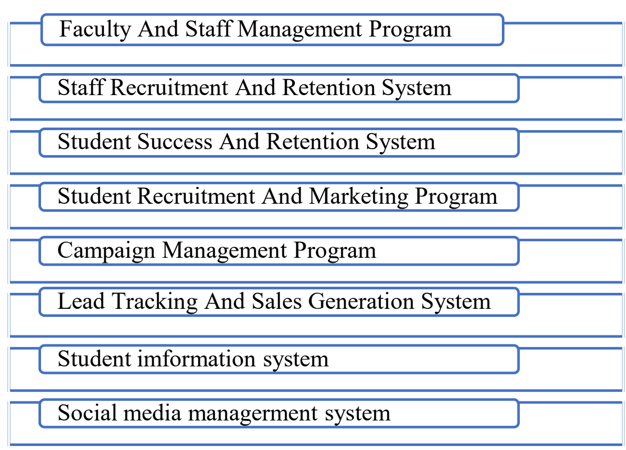


Figure 3. EliNext's CRM system solution

The author developed a process for a CRM system solution for students based on the ten features of EliNext (EliNext, 2021) and the admissions consulting and customer support process at the ICH is structured as follows (Figure 4):

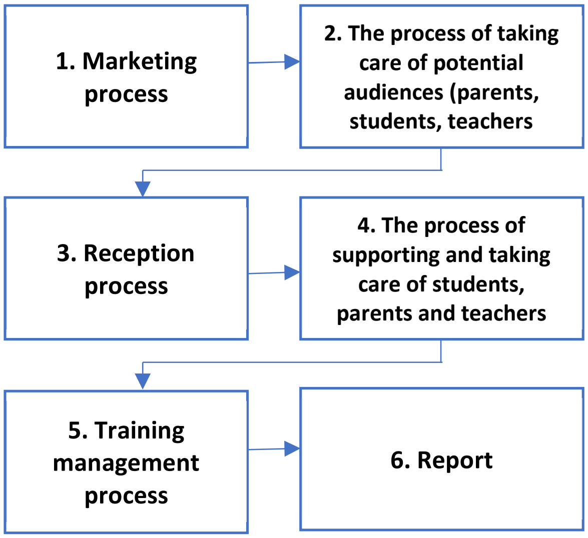


Figure 4: The proposed CRM system process at the ICH

Environment settings	
Framework:	Laravel
Server:	Linux (Shared / VPS / Dedicated)
PHP:	version 7.0 and above
MySQL Version	8.0 and above
Web server:	Nginx or Apache
Extension:	BCMath PHP Extension Ctype PHP Extension Fileinfo PHP extension JSON PHP Extension Mbstring PHP Extension OpenSSL PHP Extension PDO PHP Extension Tokenizer PHP Extension XML PHP Extension Fileinfo PHP Extensio

Figure 5. Environment settings

6.1. Marketing Process

6.1.1. Internet Communication: SEO, GG keyword, Facebook add, zalo page…, Press activities (source 1).

6.1.2. Career teaching in middle and high schools (source 2)

- Management of training programs
- Manage and organize classes
- Manage class schedule and teaching schedule
- Manage teachers and experts
- Design forms to collect reasonable information(students, parents of your students, teacher)

6.2. Process of caring for potential audiences (parents, students, teachers)

6.2.1. Aggregate data from sources by sending SMS, zalo, facebook, messenger, calling, scheduling meetings, and sending emails.

- Send e-cards for birthdays and holidays to teachers, parents, and students.
- Distribute enrollment notices to teachers, parents, and prospective students.
- Preferential vouchers for short-term and skills-based programs

6.2.2. Organize the list of prospective students.

- Private information
- Interesting topics
- Course of study (short term, long term)
- Economic circumstances
- Previous level academic results
- Do you get information from any source?
- Which of the branches is the most convenient for admission?
- Who is the consultant?

6.2.3. What stage of the process has been consulted: Consulting status; Contact count; Evaluate potential.; The reason for the consultation's failure.

6.3. Reception process

- Registration
- Examination and admission in accordance with regulations during the reception process. Plan a reception for admissions (reviews → Obtain valid admission and testing records → Disseminate the admission notice and announce the results → After paying tuition fees, transfer potential student data to official students).

6.4. The process of supporting and taking care of students, parents, and teachers

- Receiving one-time requests
- Send birthday and holiday cards to teachers, parents, and students using Zalo, the messenger app.
- Tuition fee notice to parents and students
- Send vouchers for preferential tuition fee programs for University courses and topics to lecturers, parents, and students.

7. CRM solutions and achievements contents

As previously stated, the International College's previous CRM system lacked the functionality required for effective customer relationship management. The CRM system provided by the author provides more effective methods for communicating with customers and assisting the admissions consulting department in improving work efficiency. On the other hand, the CRM system facilitates communication between the admissions department and other departments in order to reduce the number of students who drop out of the ICH.

Figure 6. Contents of CRM solution

7.1. Marketing Management:

The system offers marketing campaign management, career guidance, newspaper enrollment advice, and internet communication to collect potential leads and self-assimilate the receiving and analyzing process. To send information to customers in a proactive manner, categorize their needs. Marketing management entails the following tasks:

- *Campaign Management:* The Campaigns module manages outbound marketing campaigns and sends email, SMS, and Zalo marketing campaigns to customers (Figure 7)
- *Marketing and integration across multiple channels:* Providing information to customers through multiple channels: Zalo app..... Determine Lead revenue details, and support detailed marketing performance reports by: source, campaign, and Ni dung.
- *Report on Marketing Effectiveness Analysis:* Create a campaign budget report based on the analysis and export the excel file (Figure 8).

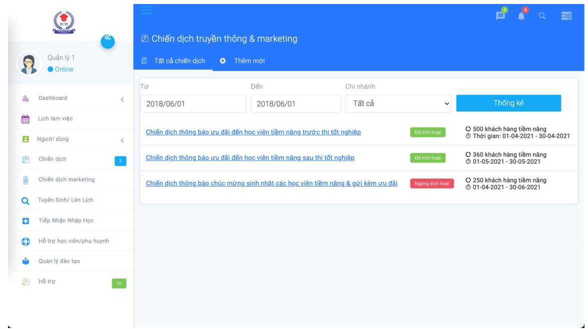


Figure 7. Module Campaign Management

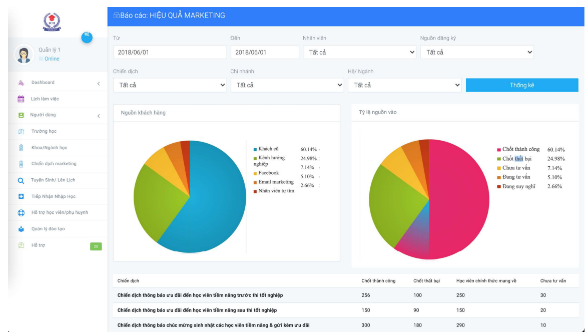


Figure 8. Report on Marketing Effectiveness Analysis

7.2. Sales management - admission consultant

- *Manage customer data:* Including raw customers (target), potential customers (lead), organizational customers (businesses), and individual customers (Figure 9**).**
- *Daily Sales Activities:* Employees plan their own work and keep track of their schedules in their accounts. Managers can quickly check each employee's schedule and work progress, and notify employees of new assignments (Figure 10, 11,12).
- *Mobile App for Sale*: Integrated mobile app for easy work management.
- *Sales Kit Sales Support:* Supports all employee lookup features such as: Research policies related to the study program, research program information and research customer problems or questions and how to handle them.
- *Report on consulting results:* Report on consulting results and employee consulting status, as well as sales closing and consulting in branches.

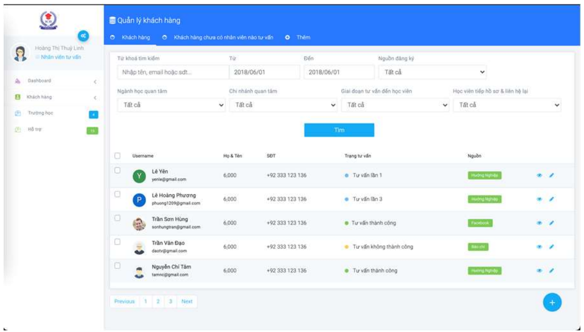


Figure 9. Customer data

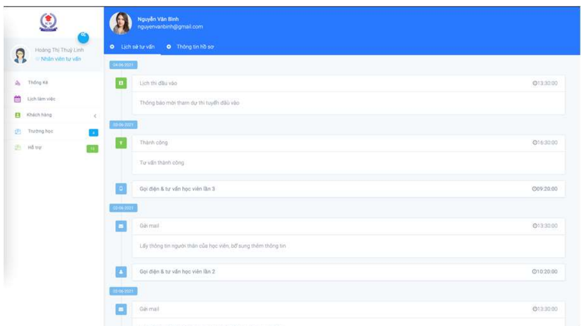


Figure 10. Consulting history

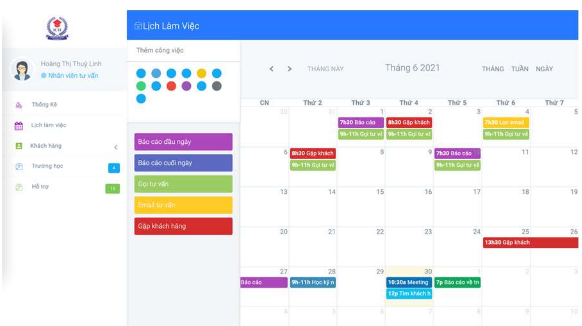


Figure 11. Employee's schedule

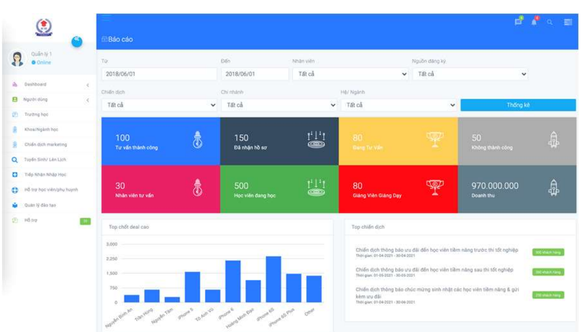


Figure 12. Efficient working report.

The CRM system's features have been evaluated as a great support tool for the Admissions Department's staff. These are reflected in the 35% increase in the number of students enrolled in the 2018-2019 academic year (Le,2019).

7.3. Customer care management

After receiving leads, staff will take care of potential customers as well as customers who are currently using the service; job functions include: Call center for customer service; SMS notification to remind customers; Zalo's customer reminder notification; Use a chatbot/messenger to nurture customers; Email/automation can be used to nurture customers; Manage customer request tickets through the portal; Administrate FAQ – knowledge repository; Customer service support report. The use of CRM has aided in better customer service; the dropout rate has decreased by 10% in school year (2018-2019) (Le, 2019).

7.4. Management of training:

The customer service department will transfer the relevant sticker to the customer's request for training staff to handle. Furthermore, the system feature will allow for the scheduling of lectures, as well as the teaching schedules of teachers and lecturers, in order to avoid duplication. Among the job duties are: Management of Scores; Management of Attendance; Organize and arrange teachers of career guidance in middle and high schools; Organize the scheduling of teachers to teach at the school.

ICH has received positive feedback from parents and students after implementing CRM to manage training.

7.5. Administration of the system:

A CRM user authorization system with a hierarchical structure, specifically as follows: User Administration; Management of Role Trees; Permission to share information; Section on Access Rights. The management board and department heads were pleased with the system's decentralized feature because it allowed them to quickly and efficiently check and assess the staff's situation.

8. Conclusion

The CRM system is not new to the industry, and it has also been shown to be effective in customer relationship management in education via research and practical applications. The CRM system developed in this study focuses on admissions consulting and customer care support (find-lock-keep), evaluates the effectiveness of marketing strategies to adjust appropriate finance, and increases student and parent satisfaction with school services. Because of the benefits listed above, the CRM system will promote maximum efficiency when used in private schools. The authors' CRM system research was implemented at the College of Economics and yielded very positive results. Specifically, increasing the number of students enrolled in the 2018-2019 school year by 35% and lowering the dropout rate by

10% compared to the same period last year (Le, 2019). However, due to the impact of the Covid-19 epidemic, the effectiveness of the current CRM system evaluation has not been assessed for the school years (2019-2020) and (2020-2021). The author's CRM system was constructed. The next direction the author team will study important variables in the CRM system when building for universities. Built and designed to optionally add or remove modules in accordance with the elements of the schools, the next direction the author team will study important variables in the CRM system when building for universities.

Reference

AACRAO. (2014). *State of CRM Use in Higher Education Report.* New York: The American Association of Collegiate Registrars and Admissions Officers.

Abubakar, M., Mukhtar, S. (2015). Relationship Marketing, Long Term Orientation and Customer Loyalty in Higher Education. *Mediterranean Journal of Social Sciences*, 46-474.

Abdullah, F. (2006). The development of HEDPERF: a new measuring instrument of service quality for the higher education sector. *International Journal of Consumer Studies*, 569-581.

Adikaram, CH. (2016). The Relationship between Customer Relationship Management and Customer Satisfaction. *International Journal of Arts and Commerce*, 69-95.

Croteau, A. and Li, P. (2003). Critical success factors of CRM technological initiatives. *Canadian Journal of Administrative Sciences*, 21-34.

Elinext. (2021, May 1st). *education*. Retrieved from Elinext: https://www.elinext.com/industries/education/

Lam, N. Q. (2015). *Research on factors influencing student satisfaction with the quality of vocational college training services at International College.* Ho Chi Minh: University of Finance - Marketing.

Le, N. T. (2019). *Report on the Results of the 2018-2019 School Year at the International College.* Ho Chi Minh: International College.

Nunamake, J. C. (1991). System Development In Information Systems Research. *Journal Of Management Information Systems* , 89-106.

Rigo, G. P. (2016). CRM Adoption in a Higher Education Institution. *Journal of Information Systems and Technology Management*, 45-60.

Stair, R. &. (2010). *Principles of imfomation systems.* Florida: Florida State University.

Wahab, S. J. (2011). he influence of usability and enjoyment on electronic customer relationship management performance in Jordan mobile communication services. *Africa Journal of Business Management*, 128-134.

Zani, A. Y. (2013). The increasing Of Students' Satisfaction and Loyaty by the Use of Customer Relationship Management. *IOSR Journal of business and Management*, 28-34.

INNOVATIVE IMPROVEMENT OF SAMPLE ORDER TOOL FOR INTERNAL SYSTEM - CASE OF SCANCOM COMPANY

Nguyen Quynh Phuong[1], Tran Thi Mong Tuyen[3] and Vo Trong Cang [2]

[1]Binh Duong University, Binh Duong, Viet Nam, nqphuong@bdu.edu.vn

[2] IMALOG, Binh Duong University, Binh Duong, Viet Nam, vtcang@bdu.edu.vn

[3]Scancom Vietnam Limited Company, Binh Duong, Viet Nam, tran.thi.mong.tuyen@scancom.net

Abstract. The account executive have spent too much time to transfer orders to Planning Department casing other parts in their work almost were not finished on time. The study presented solution of innovative Improvement of Sample Order Tool for Scancom's Internal system. In order to improve the process and saving of time working on internal sample system. Study has been performed to analyze and also calculate the average length of time for staffs when transferring orders on the internal sample system to Planning Department. The innovative process optimized steps and work time on internal sample system. In particular, was cut down redundant steps and was saved time up to 44% the working time on internal sample system for account executive. At the same time solving the delay in other parts of their routines of work. The optimization of time and work processes on internal sample system is premise in order to improve work processes in many departments such as: BOM, Warehouse, Photo, Test Center, Packing.

Keywords: Innovative Improvement; Sample Order Tool; Internal sample system; The optimization; Scancom.

1. Introduction

Enterprise resource planning (ERP) is a process used by companies to manage and integrate the important parts of their businesses. Many ERP software applications are important to companies because they help them implement resource planning by integrating all of the processes needed to run their companies with a single system. An ERP software system can also integrate planning, purchasing inventory, sales, marketing, finance, human resources, and more (Dương, 2003). Lean manufacturing is a system that includes tools and methods to eliminate wastes and irrationalities in the production process in order to reduce production costs, thereby improving the competitiveness of products in the market (Womack et al., 2007). The application of Lean techniques to the ERP system has created close connections and improved the old ordering system to spur the system's productivity to a higher level - case of Scancom company.

In the current period, facing the great global challenge is the Covid pandemic, Scancom company kind of have to cut staff but at the same time ensuring the same work performance; therefore, a matter of fact has been created a lot of challenge for every employee especially Scancom's top managers and managers in department. A challenge of sale department is the remaining employees didn't have enough time to meet the current job, causing delay order. The consequences is affected the work performance of sale department, the order has been canceled as the time for completing order can not meet the customer's demand. Scancom is a B2B manufacture so just only one sample orders that can not deliver on time for the customer meeting so that means that order with ten of thousands even hundreds of thousands of pieces have been canceled causing heavy losses for company. Therefore, deal with pressing matters, Author et al. has developed sample order system at Scancom company. Facing the challenge is the number of employees that have been cut cannot be added - limited human resources, the author et al. have focused on work process include related work to order task and sample order system. After analyzing, author found that there are some major issues in the way employees work and the incompleteness of sample order system. In there, employees spent too much time doing things that didn't go through the system such as sending emails between departments to notify about order infomation they had requested, go through an old system for ordering.

Some research has been published is Enterprise resource planning (Dương, 2003). Concepts in Enterprise Resource Planning (Brady, 2001). Enterprise Resource Planning (Summner, 2005) Lean is a dynamic process of change driven by a systematic set of principles and best practices aimed at continuous improvement. Lean manufacturing combines the best features of both mass and craft production (Womack et al, 2007).

The research aims to optimize the working steps and working time on the sample order system and dealing with delays in the rest of the employee's day-to-day work. This research is of key importance to sale department and company generally, in providing a more complete working system and solving the company's current pressing problem.

2. Literature Review

2.1. Enterprise resource planning systems (ERP)

An enterprise resource planning system is designed as a defined data structure (schema) that has a common database. This helps ensure that the information used in the enterprise is standardized and based on common definitions and user experiences. The core structures are interconnected with business processes driven by workflow among business departments (e.g. finance, human resources, engineering, marketing, operations), connecting systems and the people who use them. Basically, enterprise resource planning (ERP) is are means to integrate people, processes, and technology in a modern industrial organization (Dương, 2003).

2.2. Basic model of ERP system

The data model shows how business planning (Sales planning, and Production planning) leads detailed plans to coordinate supply chain activities. The main feature base on the common data base and their function in the system (Dương, 2003). Figure 1 shows the Basic model of ERP.

2.3. Sales Order Processing

A key part of sales order processing – accepting customer orders – is to collect current demand in kind of order forms to enter the electronic system. There are too many steps include configuring a product according to the customer's ideas, promising delivery and monitoring the status of the order. It may also include quotations - especially for custom products and outsourced through multiple factories. Designation of an order can affect demand for an item and supply chain operations. For instance, an order can be assigned to cancel or hold or to be added to the forecast, it includes an allocation, shipment authorization or an order directly linked to a resupply order. An entered purchase order can trigger communication with the actions required for other functional areas, including distribution (for dispatching), manufacturing (to assemble finished products). stocking), purchasing (ordering supplies), on-site service (to perform after-sales service tasks) and customer service (reviewing order retention) (Dương, 2003). Other functions of the basic model are presented in the reference Author et al. only mentions the necessary functions for the work.

2.4. Theoretical foundations of process analysis in Lean Manufaturing

This is one of the basic analysis methods that allows to grasp the actual situation of division of production work activities. Process analysis is one of the effective methods to detect waste and help realize improvement of each process. The purpose of this process is to clearly define the sequence of implementation stages, continue to improve in each stage, clearly define the production method, ensure the basic information for production design, basic information basis for implementing improvements, ensuring information is available for controlling production schedule (Womack et al. , 2007). In this study, authors combine Lean process analysis theory into the Sales order processing system of enterprise management software (ERP) to create a new integrated and complete system.

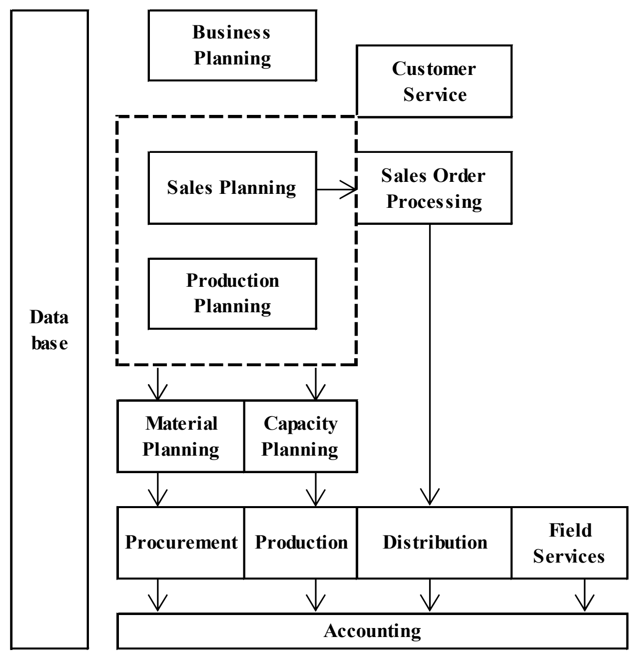


Figure 1. Basic model of enterprise resource planning system (Duong, 2003).

2.4.1. Research model:

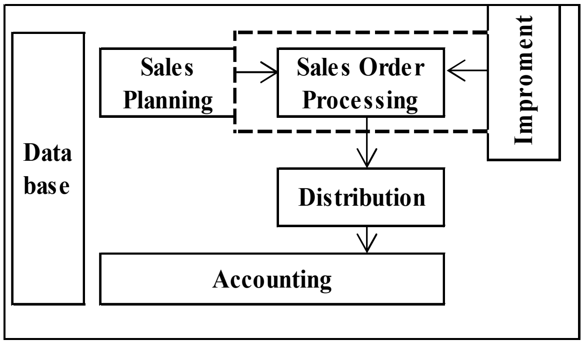


Figure 2. Improvement of model of Sample order system (Duong, 2003).

2.4.2. Improvement:

Improvement: The process is implemented in the code form on the software system to improve and perfect the sample order system.

3. Methodology

The authors measured the working time on the order sample processing system of the sales department staff and the working time related to the processing of order input. Time measurement was performed on thirty employees of the sales department. These thirty staffs are the number of retained staffs after the company's reduction due to the epidemic, most of these staffs have a relatively long working time of 3 years and above, the highest seniority is 10 year. Therefore, they are very proficient in operations on the sample order processing system, thereby minimizing the excess time caused by the incomplete operation, eliminating factors affecting the consistency of the measurement process and analysis of the study. The number of orders each employee needs to process in a day from 10 to 15 orders. The research method used in this case is empirical method, sampling time directly from employees for a working day. Sampling is carried out continuously within 30 working days of the company. The results have shown the average working time on the system and related order input has presented Table 1 below. Based on the statistics of the average working time of a staff in thirty working days, author et al. collected that an employee takes an average of nearly 4 hours to complete the input data and process information. The remaining working time of each day of each staff was spent on tasks such as meeting with customers, handling issues related to the progress with production, working by mail with customers and other tasks assigned by managers such as database statistics, reports … were not enough to meet the worked demand in routine day, leading to late orders, some changes of customers were not updated to the production... After measuring the spending time in the input data and process information, it accounts for almost half of the work time for whole day.

Table 1. The average working time of an employee within 30 working days

No. Date	Average Time of entering data (hours)	Average Time of related order (hours)	Total average time (hours)
1	1.64	1.47	3.11
2	1.63	1.51	3.14
3	1.45	1.68	3.13
4	1.43	1.69	3.12
5	1.36	1.77	3.13
6	1.74	1.41	3.15
7	1.42	1.71	3.13
8	1.39	1.73	3.12
9	1.86	1.23	3.09
10	1.19	1.9	3.09
11	1.51	1.66	3.17
12	1.4	1.75	3.15
13	1.72	1.39	3.11
14	1.56	1.52	3.08
15	1.51	1.64	3.15
16	1.41	1.72	3.13
17	1.13	1.93	3.06
18	1.79	1.31	3.1
19	1.37	1.74	3.11
20	1.26	1.83	3.09
21	1.68	1.46	3.14
22	1.73	1.29	3.02
23	1.46	1.76	3.22
24	1.49	1.63	3.12
25	1.55	1.56	3.11
26	1.24	1.92	3.16
27	1.53	1.65	3.18
28	1.49	1.62	3.11
29	1.58	1.57	3.15
30	1.59	1.59	3.18

3.1. Issues

The author analyzed the process of handling all employee input data, the author found that employees spend too much time on the following issues:

3.1.1. Issue 1:

Sales staff sent an email to the warehouse department asking to check the inventory status of the products needed for the order as well as the quality of products in stock, the feedback time for a email was not clearly specified leading the postponement in this steps and sale staff have to push email frequently.

3.1.2. Issue 2:

Sale staff checked the available quantity and product quality reported from the warehouse's mail and adjusted the quantity for ordering accordingly.

3.1.3. Issue 3:

Sale staff did repetitive operation of entering data into the system for the same products of each order.

3.1.4. Issue 4:

Sale staff worked outside the system by emailing Planning about additional requests in special situations of the order.

3.1.5. Issue 5:

Sale staff entered order data into another system because the main system hadn't have this function, Planning department must put together information and respond via email then sale staff input the status of the order into the main system.

3.1.6. Issue 6:

Sales staff send order emails to the outsourcing department, the feedback time for a email was not clearly specified leading the postponement in this steps and sale staff have to push email frequently.

3.1.7. Issue 7:

Sale staff sent email asking for price for each product in each order, in many cases there are special requirements for each product price, the feedback time for a email was not clearly specified leading the postponement in this steps and sale staff have to push email frequently.

3.1.8. Issue 8:

Order reports and many reports is relative order information have to collect from many different sources and not fully updated from other individuals, making it difficult and time consuming to report the work.

3.2. Solutions

Authors propose the following solutions to solve the current issue has been shown as below:

3.2.1. Solution 1:

Author et al. proposes to set up an additional checking section in the order window on the sample order system, set up a new window for the warehouse department to operate on the sample order system, when the sale staff enters order into the sample order system and clicks on the item to check inventory of products, the warehouse department immediately receives information of orders from the system then they check and update the status of products on the system within 30 minutes after receiving the required information from the order system.

3.2.2. Solution 2:

Author et al. proposes to create a formula under code form on sample order system, after receiving the updated product quantity information from the warehouse department, the system automatically calculate the remaining quantity in stock and automatically announce the quantity need to order to the planning department.

3.2.3. Solution 3:

Author et al. proposes to set up a new window in the system to integrate the excel standard format to import the excel file into the system without entering each item into the system.

3.2.4. Solution 4:

Author et al. proposes to create note on the system in the order interface to note special requirements and add more informations about orders, so that employees work on a same window on the system avoid sending email back and forth among many parts.

3.2.5. Solution 5:

Author et al. proposes to move the sample order program on the old system to the main system so sales department and planning department can work together on one system - main system.

3.2.6. Solution 6:

Author et al. proposes to set up an additional order section in the order window on the sample order system, set up a new window for the oursousing department to operate on the sample order system, when the sale staff enters order into the sample order system and clicks on the item to purchases, the oursousing department immediately receives information of orders from the system then they feedback on the system within 30 minutes after receiving the required information from the order system.

3.2.7. Solution 7:

Author et al. proposes to add a price request click and a note of special price requirements on the same window for sales staff operate order, set up a new window for the pricing department to operate on the sample order system, then they feedback on the system within 30 minutes after receiving the required information from the order system.

3.2.8. Solution 8:

The author proposes that after all operations are performed on the same system, the collection of information to report and check the status of the order is complete and highly accurate because all departments are responsible for the work they are in charge of on the system.

4. Results

Author et al. collected information then proposing solutions based on the outstanding problems for a period of seven working days. The process of implementing solutions from natural language to programming language with programmers was implemented for a period of one month before launching the test program for sales staff to experience. The completed program is run on the test platform to observe and test the effectiveness of the system after being improved. Testing on the new platform was carried out within 7 days to check the convenience and working time on the improved system. The convenient test results are presented in Table 2 below:

Table 2. Statistics the convenience of the test system within 7 days.

Convenience	Percent
A: Absolutely no problem using it	33%
B: No problem using it	25%
C: Normal use	20%
D: Having some problems when usin	18%
E: Difficult to use	4%

Statistics show that the factor "Absolutely no problems when using" accounted for 33%, the factor "No problems when using" accounted for 25%, and the factor "Normal use" accounted for 20%. These three factors account for a total of 78%, showing that the comfort of the improved system is highly appreciated. The factor “Having some problems when using” accounted for 18%, and the factor “Difficult to use” accounted for 4%. This total percentage is 22%, the author and manager have recorded it and taken appropriate adjustment measures after consulting the sales staff - the main operator on the system. The measuring result of the working time on the improved system is presented in Table 3 below:

Table 3. Statistics of working time on the test system within 7 days

Date	Day 1	Day 2	Day 3	
Average time of an employee (hours)	2.15	2.05	2.03	
	Day 4	Day 5	Day 6	Day 7
	2.09	2.07	2.17	2.07

The statistics show that the order processing time on the sample order system of an employee was from 2.03h to 2.17h within 7 days. Compared to the average order processing time on the system for an employee before improving system had from 3.01h to 3.19h. Therefore, the time difference between two parts was from 0.98h and 1.02h. it was equivalent to a 32 % reduction in the total average working time on the system of sale staff.

After making statistics within 7 days, Author et al. proceed to change the outstanding issues in construction of the ordering system (these issues have been presented in the discussion and recommendation of the paper) in order to implement the improved sample order system in real work. After editing and calibration within one week, the new system came into official use. Statistics on convenience and usage time on the new sample order system were made regularly every 7 working days until the end of 2 months after using the new system to ensure updating and efficiency can be adjusted promptly and quickly. Statistics of working time on the sample order system within 8 weeks as below:

Table 4. Statistics of working time on the sample order system within 8 weeks

Week	Week 1	Week 2	Week 3	Week 4
Average time of an employee (hours)	1.76	1.65	1.68	1.69
	Week 5	Week 6	Week 7	Week 8
	1.6	1.64	1.72	1.71

Table 5. Statistics the convenience of the test system within 8 weeks

Convenience	Percent
A: Absolutely no problem using it	67%
B: No problem using it	25%
C: Normal use	8%

Statistical results after 2 months of using the improved sample order system show that it is very feasible, specifically, the working time related to orders in terms of data entry and handling order issues has been reduced by 44%. The convenience of the improved sample order system completely meets the needs of the employees. And some reference images of the improved sample order system are presented in figure 3 below:

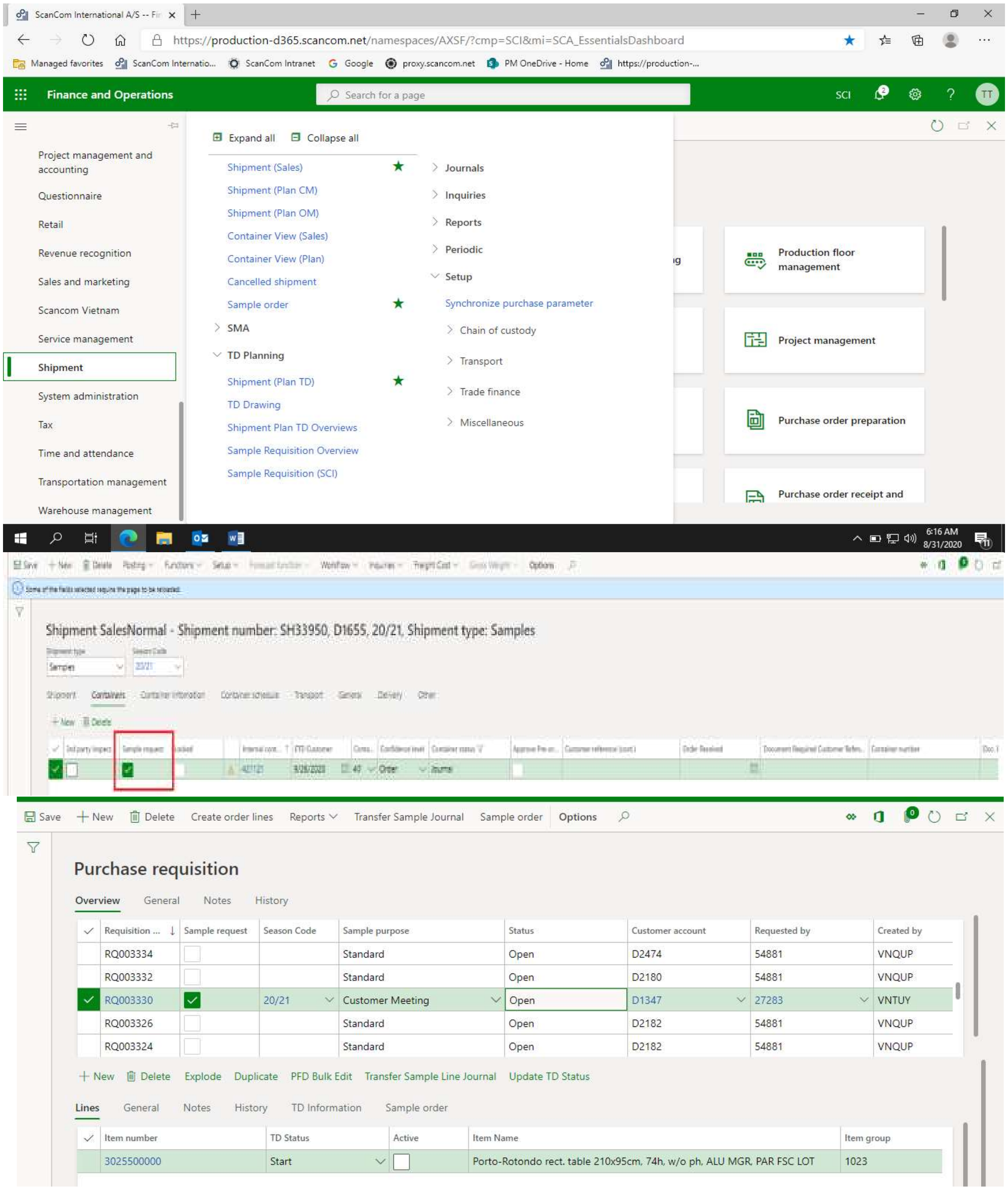


Figure 3. Some Interfaces screen-shoted from the software

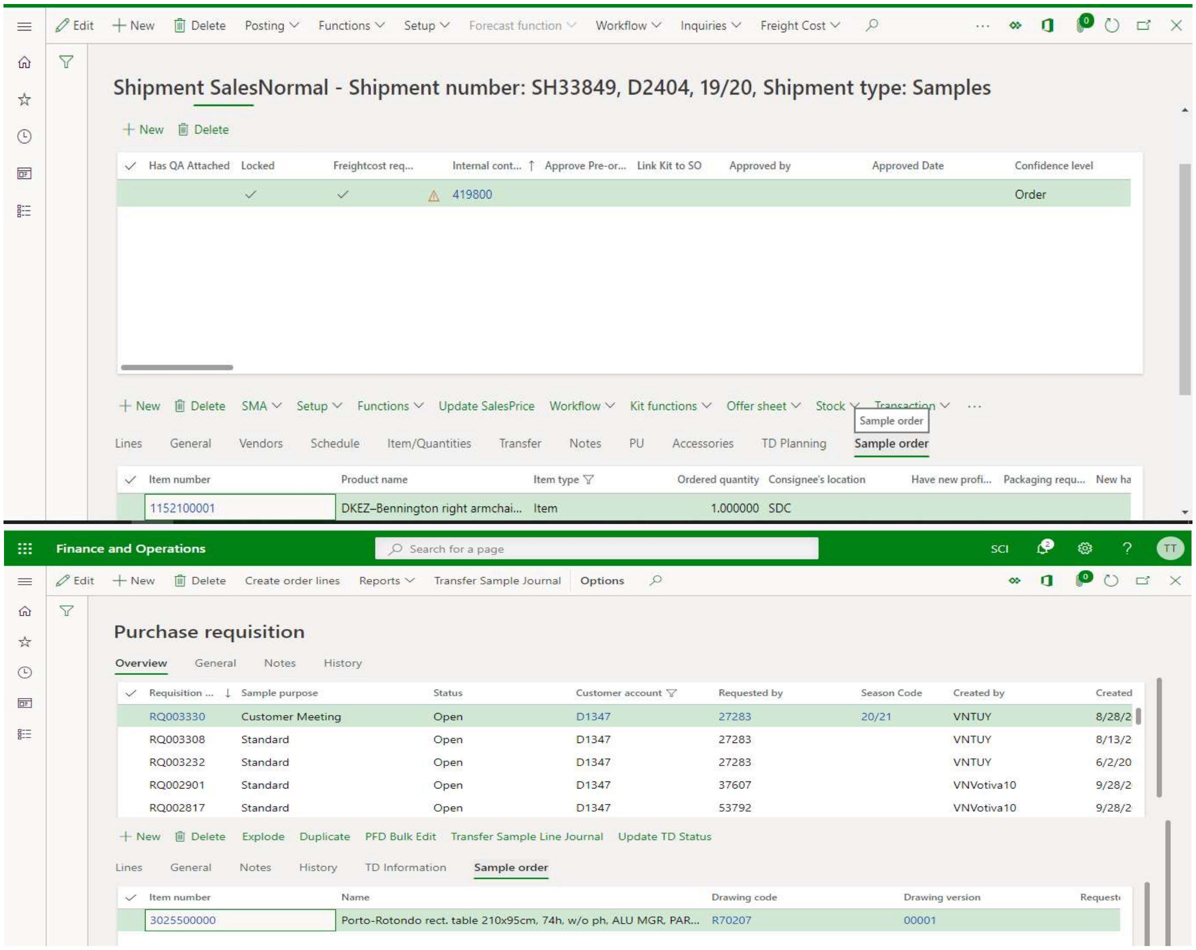


Figure 3 (cont.). Some Interfaces screen-shoted from the software

5. Discussion and Recommendation

System improvement based on the theory of enterprise resource planning (ERP) system and lean process analysis (LEAN) has given an integrated and complete ordering system beyond expectations. The study solved the company's pressing problem:

- Using the company's current limited resources effectively, improving human performance.
- Reduce working time on the system up to 44% and cut off redundant stage of the working day.
- Collecting information to report weekly, monthly and quarterly for reporting becomes extremely easy and intuitive for collectors and managers.
- Solving the company's urgent problem is to prevent the consequences of losing orders from customers.
- Complete and integrate the new system effectively to help users achieve good performance.

Using limited resources creates a great challenge for the research process and also creates a root case for research. Through the research results, Author et al realize that the waste of using human resources has existed for a long time and learn management experience in the future. The reduction in working time has been proven by research to help employees in the sales department be more smooth and fluent in the working process, helping these employees to complete their work more efficiently and scientific. The problem of collecting information to report to managers is no longer a dilemma each week, especially at the end of month. The reason that reports put a lot of pressure on employees is top managers and managers based on the information of each report to know revenue, predict output, demand for purchasing raw materials for production, etc. Due to the interconnectedness of the report, it is necessary and time consuming to provide information that needs to be accurate. Through system improvement, the system is completely improved, all users need to do is to export the excel file according to the pre-programmed report template needed for the purpose of each report for different departments. Solving the problem of losing orders for the company has brought meaning not only to overcome the loss of orders but more importantly, to help the company overcome this pandemic, to help the company in the future. be able to once again re-hire the workers they had to lay off, as well as expand the business in the near future for making a good impression on old and new clients by meeting deadlines well for orders during this pandemic. Solving the problem of losing orders for the company has brought meaning not only to overcome the loss of orders but more importantly, to help the company overcome this pandemic. It helps company be able to once again re-hire the workers they had to lay off in the future. As well as expand the business in the near future for making a good impression on old and new clients by meeting deadlines well for orders during pandemic.

Limitations of the study, as well as other studies, except for its application on limited human resources, are the problems of accuracy and comprehensiveness and the implementation time of the new system after being improved and they need to be corrected. The information entered into the system not only needs to be accurate for the employees of the sales department, but also the feedback information on the system of other departmental employees must also be accurate. Therefore, there were meetings to determine the rights and responsibilities of the parties involved when operating on the main system. The author has received feedback from the warehouse department for products that are not good enough to ship but still have components that can be reused because they are still guaranteed in quality. We recorded the information and proceeded to add a note window on the system so that the planning department can refer the information to adjust the materials accordingly. The response time from the departments according to the system regulation is 30p for the warehouse department and 24h for cost department, which caused controversy when different departments responded that there was still a lot of work to be done and it would be difficult to meet the scheduled time, not to mention the cut off staff has also happened in their departments. Facing this difficulty, Author et al. proposed meetings to find a common solution and reached an agreement for responding on system are 45 minutes for the warehouse department and 36 hours for new price, 15 minute for existing price for the cost department. As well as recommending system improvements for the warehouse department and the cost department in the near future. Producing system improvements for the warehouse department and the Cost department in the near future. In the near future, the author and manager will continue to research and improve the new system for the warehouse department and the cost department. Besides, they will learn and ask experts who have experience in system improvement and management.

Reference

Brady, J. (2012). Concepts in enterprise resource planning. Course Technology.

Dương, Quang Thiện (2003), Enterprise resource planning.

Sumner, M., & Hamers, R. (2005). Enterprise resource planning.

Womack, J., Jones, D., & Roos, D. (2007). The machine that changed the world. Free Press..

APPLICATION OF SYSTEM ENGINEERING IN DESIGN OF SOLAR ENERGY SYSTEM: A CASE STUDY

Dat Minh Le[1], Phuong-Nhi Thieu Pham[1], Quang-Hao Duong[1], Nam-Anh Diep[1]

[1]Department of Industrial Systems Engineering

Faculty of Mechanical Engneering, Ho Chi Minh City University of Technology - VNU

Email: dat.lemindast@hcmut.edu.vn, nhi.pham.ihn@hcmut.edu.vn,

hao.duong0109@hcmut.edu.vn, 1870183@hcmut.edu.vn

Abstract. The research examines the application of systems engineering approaches to utilize new system design of renewable solar energy which has been successful to develop many complex projects such as ship building project, electronic component manufacturing project… The project is also specially important meaning in case of considering the application of this technique in the design of complex systems that are directly influenced by natural factors or utilize natural factors such as in energy suppling section as a typical example. The research will apply systems engineering in designing solar energy system applied available roofs of warehouse building. It aims a completed design – layout and specification of device meeting the needs of stakeholders, natural conditions, available physical condition and all likely aspects of the systems are considered and integrated into a whole. These solutions propose an optimal, easy-to-apply design to actual electrical system designs using similar natural resources, while at the same time making advantage of the resources available to the building owner.

Keywords: solar energy, solar panels, systems engineering.

1. Introduction

Systems engineering is a multidisciplinary method used to develop, test an optimal and integrated set of product designs, processes that satisfy user demands, and provide information to have a decision. The function of systems engineering is to guide the engineering of complex systems. System engineering includes the identification of customer demands, the system's operating environment, the system's interface, logistics support requirements, the capabilities of the operators and other concurrent components. Principles of systems engineering - holism, emergent behavior, boundary, et al. - can be applied to any system, complex or otherwise, as long as systematic thinking is used at all levels.

The energy sector plays a key role in promoting the country's socio-economic development. However, fossil fuels are being depleted over time, and the negative effects of their use are becoming more severe and alarming. This is why renewable energy, or in particular solar energy, is increasingly replacing fossil fuels with the goal of a clean energy future. Solar energy is clean, enormous, infinite, ubiquitous energy that we can exploit. In recent years, countries around the world are together exploiting and putting this clean energy source into use. The extraction process does not cause any negative effects on the environment. In contrast, solar energy brings many other benefits such as creating green energy, protecting the environment, saving electricity costs, easy maintenance and long life. The mechanism of action of this device is based on the photoelectric effect in physics.

Vietnam is one of the countries with great potential to develop solar energy. Our country belongs to the subtropical monsoon region, has a coastline of more than 3,000km, the central and southern regions have a total of hours of sunshine in a year ranging from 1,400-3,000 hours. average solar radiation is about 4-5 kWh / m2 / day.

ABC Company is a foreign enterprise operating in the wood furniture sector in the demanding markets of America and Europe. The company plans to take advantage of the solar roof system to reduce electricity costs. In addition, certifications of using renewable energy can be leveraged to promote business.

Therefore, in this paper, a system engineering approach is applied to the design of rooftop solar PV systems for this company by gradual steps with illustrations and calculations that have been applied in practice to optimize and utilize available resources.

2. Literature review

In the holistic view, systems engineering focuses on analyzing and eliciting customer needs and required functionality early in the development cycle, documenting requirements, then proceeding with design synthesis and system validation while considering the complete problem, the system lifecycle. This includes fully understanding all of the stakeholders involved (Oliver et al., 1997) claim that the systems engineering process can be decomposed into: Systems Engineering Technical Process, and Systems Engineering Management Process.

Within Oliver's model, the goal of the Management Process is to organize the technical effort in the lifecycle, while the Technical Process includes some steps to develop:

- Assessing available information,
- Defining effective measures.

With the purposes to:

- Create a behavior model,
- Create a structure model, perform trade-off analysis,
- Create sequential building and testing plan.

Depending on their application, although there are several models that are used in the industry, all of them aim to identify the relation between the various stages mentioned above and incorporate feedback. A system engineering approach to building a complete system includes the following five (05) steps (Kossiakoff et al., 2011) combining together with seven (07) steps of design process of solar energy system (Thang and San, 2014) has been developed as a popular approach in the designing and constructing small and medium solar power projects in the direction of taking advantage of available facilities and installing in residential areas or workshops of industrial zones in Vietnam, will be observed to apply throughout the paper, simplified and shown as below:

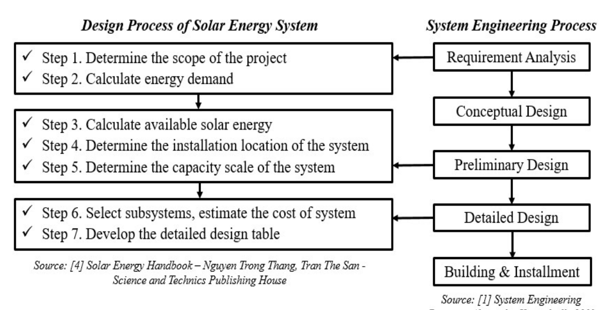


Figure 1. The combining together between system engineeting approach and (07) steps of design process of solar energy system

The practical application based on the data collected from the onsite will be presented in the following sections of the paper. This includes:

- Calculations centered on the number of photovoltaic solar panels consider infrastructure limitations (Thang and San, 2014);
- Consider the structure involved as well as the natural conditions for system development in the locality where the system is being planned to be built (Arno et al., 2015; Philipps et al., 2017).

3. Problem Statement

ABC Company is one of the leading companies in the field of wood furniture production with main premises located in Nam Binh Duong Industrial Park. Currently the company has a new finished product warehouse located nearly to improve the storage capacity. With the increasing energy demand and stringent standards on emissions followed the updated regular of parent company, ABC Company is looking for a new source of energy, considered as "cheaper" and "cleaner" to be able to maintain the warehouse operations effectively.

Based on the geographic location of the factory area, it is potential to make advantage of the amount of sunlight energy for both the internal consumption and exporting to outside trading. With an average solar energy in the year 1,621 kWh/m2/year, it is expected to generate energy that can meet the needs of warehouse operations and, moreover, the possibility of excess electricity to sell to the state load.

For the above reasons, there has been an opportunity and aslo challenge for ABC Company to develop new energy system.

4. Need Assessment

The source of information used for the research is plentiful, coming from many sources with both qualitative and quantitative nature, supplied by many stakeholders, data collection for need assessment step used 03 popular methods including: expert opinion method, onsite survey method, and direct interview method.

4.1. Solar Panels System Installation and Construction Process

The below process is based on design experience following the expert opinion method which emphasizes the importance of actual survey to make sure the theoretical system design meets the requirements of both actual conditions and stakeholder.

Figure 2. Solar panel system installation and construction process

- Step 1: Consulting customers a suitable solar power system through area, budget...
- Step 2: Experts conduct field surveys, dividing light conditions, area, and inclination angle during the construction of solar energy to contribute to the optimization of the solar system.
- Step 3: Design the solar energy system - Survey results are analyzed and systematized, then measured and calculated, financial costs and payback points are calculated.
- Step 4: Construction and installation of solar power system.
  - Choosing the right direction and angle for solar PV system installation.
  - Construction of solar cell systems installed in the top-roof.
- Step 5: Machinery supplier conducts the warranty and periodical maintenance period after the construction of solar power: Warranty the equipment of the solar power construction system in accordance with the factory standards.

4.2. Solar Panels System Operation Process

In principle, the solar power generation system can be roughly described as shown in Figure 3 according to the principle of converting energy from collectable solar radiation to alternating current that carries the ability to transmit commercially in long distances.

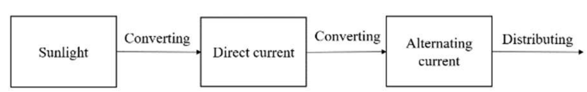


Figure 3. Solar panels system operation process

Sunlight conversion: Solar panels absorb sunlight and convert into direct current: As soon as the photovoltaic cells are exposed to sunlight, photons will be absorbed by the semiconductor materials. From there the photo-electric phenomenon occurs and the liberated electrons move freely and generate direct current. And then this current will be sent to the on-grid Inventer.

Converting DC to AC power: The inverter (the grid unit) converts DC current (one-way) from solar panels into AC (AC) current of the same phase, same frequency.

Distribution of electricity: If there is enough sunlight, the panels will produce enough electricity for the households, the system will give priority to using solar electricity. Conversely, if it is not sunny, the system will automatically compensate for the grid power. If the amount of electricity from the solar cell is not used up, it will be pushed to the grid, the electricity will install the 2-way meter for free, to measure the excess electricity from the solar panels system.

4.3. Stakeholder Survey

Based on the many survey times in the premisess, stakeholders of the system are determined including: Investor, Regional power company (National grid system), Construction department, Raw material supplier, Maintenance department. After surveying the Stakeholders, we have the following SRD (Stakeholders requirement document) table.

Table 1. Stakeholders requirement document (SRD)

Stakeholders	Requirement
Investor	• The end product meets the requirements of capacity, durability, stability, safety... • Ability to control and manage easily. • Total investment is about 30 billion. • Time of use at least 25 years. • Construction time is not more than 3 months. • The construction process does not affect warehouse operations.
Regional power company (National grid system)	• The investor needs to carry out the procedures for supplementing the solar power development planning, and the procedures for licensing electricity activities in accordance with the State's regulations. • Ensure proper working process, register to sell solar power grid. • Comply with the requirements of Circular 18/2020-BCT, Decision 13/2020-BCT.
Construction department	• Making the contract, reducing problems. • Need a complete monitoring unit to avoid scheduling overlap between units
Raw material supplier	• Make sure to agree to the agreements, purchase and sale regulations, product warranty policies.
Maintenance department	• There are periodic cleaning equipments. • Follow up construction and repair process.

4.4. Project Constraints

Followed the opinions of expert method and investor (also owner of the designed project), there are some constraints observed in the designing period of the research system are shown in the following table.

Table 2. Project constraints

Factor	Constraints
Amount of sunshine	More than 5 hours per day.
Sunshine quality	The optimal angle is 15 → 45 degrees to the south, particularly in the South 16 → 18 degrees for the maximum efficiency of the panels.
Factory area	18,000 m^2.
Time of use	25 - 35 years.
Initial investment	30 billion VND.
Power capacity	2.3 MWP / year.
Construction time	Less than 3 months.

4.5. System Requirements

Based on the data and design conditions collected and analyzed above, an aggregation is made to highlight the design requirements required for the next designing steps.

Some notes when installing solar power systems:

- ***Location:*** System is located properly then solar cells can capture sunlight, at least between 9am and 3pm. Systems such as rooftop solar cells depend a lot on a good location and the direction of the panels.
- ***The ability to match the roof:*** When installing solar panels systems on older buildings, it is important to consider whether the roof can withstand the weight of the solar panels system. The area of the roof and the possibility of construction determines the number of solar panels. Based on the size of your home, draw your roof diagram on a piece of paper, determining the specific direction. Using the free Google SketchUp tool, the AutoCAD program or contact a professional consultant can accurate and have reasonable calculations.
- ***The size of the solar panels:*** Solar panels come in all different shapes and sizes. Power demand calculation will provide a more general overview of the size of the system needed.

Some notes about technical issues: The area of the roof and the possibility of construction determines the number of solar panels. Those in charge need to calculate the amount of energy from the battery system solar energy can generate consistent with the actual needs and take note of shading issues as well as output. They also need to choose suitable solar panels, choose inverter technology for solar power installation, choose accessories for mechanical frame for solar PV installations... In addition, "cloud computing" technology can be used in the solar power monitoring system.

5. Conceptual design

5.1. Operation Analysis and System Function

Deploying the detailed design from the current conversion principle diagram as shown in Figure 3, we have a general device distribution diagram presented as below Figure 4 and the explanation of functions of the important parts.

Figure 4. System operation

- Sunlight transformation: solar panel systems absorb sunlight and convert it into one-way currents. And then this current will be sent to the inverter (grid).
- Converting one-way electricity to AC: The inverter (the grid-connected unit) converts DC current (one-way) from solar panels into AC (alternating-current) current of the same phase and frequency.
- Control function: the system of switches connected in series with the system of lightning surge fuses, ensuring current interruptions when problems occur or in cases of active disconnection for maintenance and repair. In addition, there are also screens that display system voltage parameters, amperage of the current system. The alternating current is then fed to the substation and connected to the national grid.

Based on the stakeholder requirements summerised as below Table 3, the design team gave a number of directions to apply to the project. These are the basis for the design team to continue developing the solar power system.

Table 3. Design direction based on the stakeholder's requirements

Stakeholders	Requirement	Design direction
Investor	- The product system meets the requirements of capacity, durability, stability, safety... - Easy to control and manage. - Total investment come from development budget of company is limited within about 30 billion (about 1,300,000 USD) for new energy system. - Life span: at least 25 years. - Construction time: not more than 3 months. - The construction process does not affect warehouse operations.	- Reasonable, economical design. - Take advantage of available resources to save investment and operating costs. - Choose a reasonable installation time, not affected by weather and do not affect the operation of the below warehouse. - Ensure the correct construction schedule.
National Grid	- The investor needs to carry out the procedures for supplementing the solar power development planning, and the procedures for licensing electricity activities in accordance with the State's regulations. - Ensure proper working process, register to sell solar power grid. - Comply with national standards.	- Make sure that the system meets technical standards for electricity of Vietnam Electricity (EVN).
Construction department	- Making the contract, reducing problems. - Need a complete monitoring unit to avoid scheduling overlap between units.	- Make optimal construction plan. - Comply with the construction safety regualtions and arrange additional supervisors at the onsite to supervise.
Raw material supplier	- Make sure to agree to the agreements, purchase and sale regulations, product warranty policies.	- Depending on the calculations, select the contract according to detailed cost estimates or package contracts.
Maintenance department	- There is periodic cleaning equipment. - Follow up construction and repair process.	- Cooperate with reputable raw material suppliers - Selected products must ensure the standards of design, installation ...

The following figure shows the main functions of the solar energy system has been detailed and developed (shown as below Figure 5) based on the function of each component.

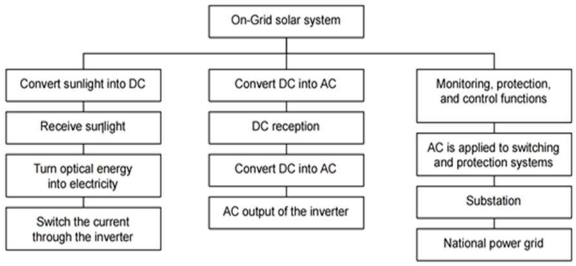


Figure 5. Diagram of system functions

In general, the solar power system has 3 main functions. Depending on detailed function, the design team has broken down the system to each specific sub-function. From these subsystem functions, the design team will come up with ideas and designs to perform these required functions.

5.2. Feasibility Study

5.2.1. Natural Condition

Southern Vietnam is sometimes sunny all year round. In January, March, and April, it is usually sunny from 7:00 a.m. to 5:00 p.m, average radiation intensity is usually greater than 3,489 kWh / m2 / day. Binh Duong has a favorable geographical location, the solar energy potential is considered good compared to the national average. The number of sunny hours in the year in the province ranges from 2,200 to 2,800 hours, with the average solar radiation of about 4.5 kWh / m2 / day.

5.2.2. Feasibility Analysis of Grid Connection

The following table show the feasibility analysis of grid connection supplied by the local supplier shown the load capacity at the local power grib system in the present time conducted the research throughout the local transformer stations.

Table 4. Feasibility analysis of grid connection

Name of TBA /line	**Rated capacity of TBA / line (kVA)**	**In-operation rooftop solar PV capacity (kW)**	**Rooftop solar power capacity with connection capacity (kW)**
Khanh Binh Transformer Station			1,128
Khanh Binh T1 Transformer Station	1,581	0	458
Khanh Binh T2 Transformer Station	1,581	0	458

Conclusion: In case of there is no plan to upgrade the substation in near future conducted by local supplier, the capacity of the system must be reduced compared to the original estimate. If the system capacity is greater than the available capacity of the substation, and the station is not upgradeable, the capacity of the original system can be reduced.

5.2.3. Force-Resistance of The Structure

Process of calculating force-resistance of the structure based on the master layout (shown as in Figure 6 of the premise with 05 steps of design process:

- Step 1: Collect construction drawing data and actual measurement.
- Step 2: Live load analysis.
- Step 3: Set up parameters on the calculation software.
- Step 4: Check the ability to withstand structural forces, bolt connections, welds ...
- Step 5: Upgrade and reinforce the structure if necessary.

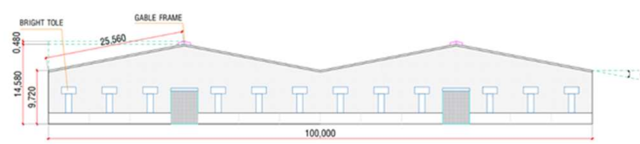
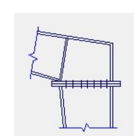
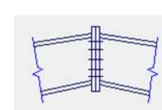


Figure 6. Warehouse data collecting

Table 5. Truss frame structure and types of connections in structures

Truss frame structure			
Frame distance:	10m	Beams-column:	
Number of frame steps:	20		
Purlin type:	C20		
Purlin distance: (m)	1.4m	Top beams:	
Large beams for column links: (mm)	250 x (1200 x 500) x 12 x 10		
Connecting beams to the roof: (mm)	250 x (500 x 350) x 12 x 10		
Boundary colum: (mm)	250 x (1200 x 500) x 12 x 10	Top beam between two roofs:	

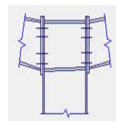
Middle column: (mm)	250 x 600 x 12 x 10		
Material:	Steel CCT42		
Bolts:	2 chains (8 each side)		

Table 6. Live load analysis

Load analysis		
Static load: 10 kg/m^2	10x9 = 90kg/m = 0.9kn/m	Loads of toles, purlin
Live load: 30 kg/m^2	30x9 = 270kg/m = 2.7kn/m	Live load repairing, installation
Solar PV load: 13 kg/m^2	13x9=117kg/m=1.17kn/m	Load capacity of the solar energy system

Calculating the force-resistance of the structure

The calculation of the load indexes of the construaction works is often normative and uniformly applied in the construction engineering community, simplifying the calculation through specialized software, considered the parameters as input data and calculate matching results. The calculation is not considered as main reseach subject in the paper, but throughout which to demonstrate the suitability of the design to the existing construction conditions. The diagram below shows the important metrics and steps required to enter computational data for the software.

Using KIW pre-engineered steel frame calculation software KIW has been developed as CIC - one of famous brand on construction community in Vietnam, specialized in loading and designing framework construction - is an automated software to analyze the design of prefabricated steel frame structures, giving the final results as construction technical drawings for steel frames with I-shaped bar components. In particular, KIW is capable of the optimal design ensures the most impact-resistant and material-saving conditions.

As shown in Figure 7 is a step-by-step guide to inputting construction data for the building, issued 2D simulation layout shown as Figure 8a. After checking the forces for positions, output results in form of Excel file as shown in Figure 8b, allowing preliminary assessment based steel frame structure data of the warehouse.

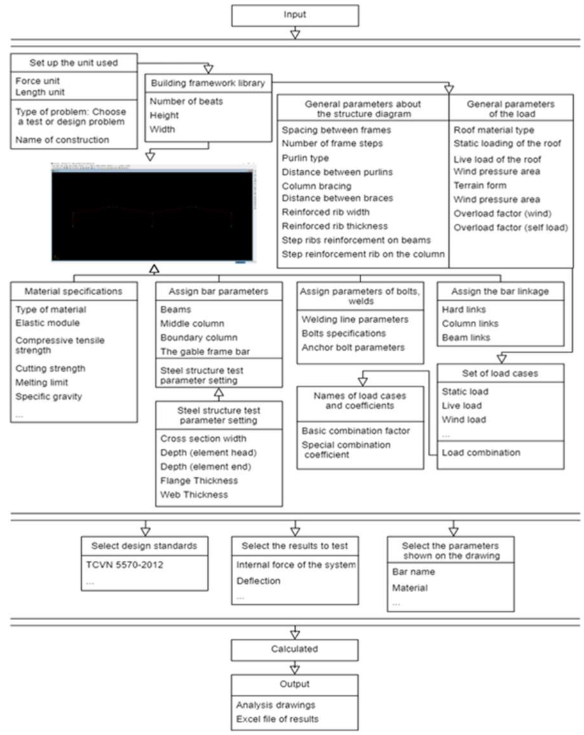


Figure 7. Import and export operations and work with the KIW software

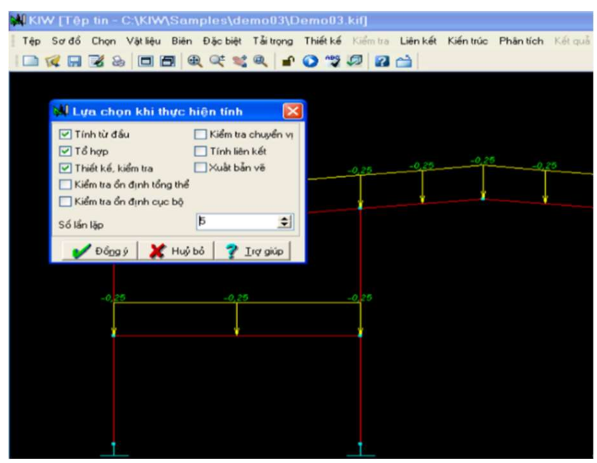


Figure 8a. KIW software - User interface

Phần tử	Mặt cắt	Tiết diện (mm)	KNCL	ODTT	ODBB	ODBC
1	1	Ith(900x250x10x12)	Đạt	Đạt	Đạt	Đạt
	2	Ith(800x250x10x12)	Đạt	Đạt	Đạt	Đạt
	3	Ith(700x250x10x12)	Đạt	Đạt	Đạt	Đạt
	4	Ith(600x250x10x12)	Đạt	Đạt	Đạt	Đạt
	5	Ith(500x250x10x12)	Đạt	Đạt	Đạt	Đạt
2	6	Ith(350x250x10x12)	Đạt	Đạt	Đạt	Đạt
	7	Ith(384x250x10x12)	Đạt	Đạt	Đạt	Đạt
	8	Ith(425x250x10x12)	Đạt	Đạt	Đạt	Đạt
	9	Ith(466x250x10x12)	Đạt	Đạt	Đạt	Đạt
	10	Ith(500x250x10x12)	Đạt	Đạt	Đạt	Đạt

Figure 8b. Corresponding table of result

The inspection will be carried out for columns, beams and overall deflection of the steel frame, including 06 result tables as shown in Figure 8b. This combination of result tables gives a comprehensive assessment of the steel-frame structure of the building, in role of important feasibility conditions for system installation.

After the feasibility calculation process, the structure of ABC workshops - 180x100m warehouse has enough force-resistance capacity to develop the solar panel system, demonstrating the feasibility of the project in terms of existing steel-frame structures.

Trade-off analysis has been conducted and shown as below Table 7:

Table 7. Select the type of system suitable for the purpose of use

Criteria	Grid-Tied Solar Systems	Off-grid Solar Systems	Hybrid Systems
Utilizing excess electricity	Yes Unlimited storage	No Limited storage	Yes Unlimited storage
Cost	- No storage batteries cost, much cheaper than Off-Grid system.	- Costs of storage batteries and power banks (up to 40% of project costs)	- The highest cost in 3 options
Backup power for power cut	No	Yes	Yes
Process and procedure of operation	- Registra-tion for grid connection, sale of electricity, testing procedures for capacity standards, installation, safety... at the request of the power company	- Self-sufficient, no strings attached.	- Registra-tion for grid connection, sale of electricity, testing procedures for capacity standards, installation, safety... at the request of the power company
Using characteristics	- Capable of gaining capital and making a profit after 4-5 years. - Suitable for large scale solar power projects.	- Independent from the grid. - Autonomous in power source. - Suitable for household scale projects, areas prone to power cut	- Capable of gaining capital and making a profit after 4-5 years. Suitable for areas prone to power cut. - Suitable for large-scale solar power projects.

Based on the characteristics of ABC Warehouse: Located in Nam Tan Uyen Industrial Park, the regional power system is always stable.Stakeholder functional analysis: Large solar power system project scale (over 30 billion), large installation area, with system purpose of selling electricity... Design team decided to choose the first option: Solar power is connected to the grid to take advantage of the infinite storage capacity of the grid while minimizing costs.

5.3. Risk Assessment

The risk assessment is mentioned basically some hazards should be considered in the constrcution and machinery installment period followed the expert opinions is shown as in below Table 8.

Table 8. Risk assessment board

Risk	Severity	Frequency	Remedies
Construction accident	Serious	Average	- Select reputable construction contractor - Strict monitoring of safety standards
Structure collapse	Serious	Low	- Use standard materials - Timely maintenance
Reduced capacity	Medium	Average	- Timely maintenance
Bad weather, natural disasters	Low	Average	- Have an automatic generator

5.4. Functional Baseline

In above steps, the research has been considered the requirements of the stakeholders, after performing the analysis steps, reviewed the current condition of workshop structure by specialized software with suitable equal. Some basic risk assessments and operational scenarios also have been developed to ensure the new solar system design can work effectively and suitably together with the power local infrastructure.

Depended on the collected data, the warehouse system is developed consisting of 06 main subsystem, there are 03 subsystems of them including: Physical, Safety and Interface, will be emphasized to develop detailed and shown as below.

Functional baseline was built for the warehouse system shown as below Figure 9, and then system requirements specification table - Table 9 has been totally implemented for next steps of design process.

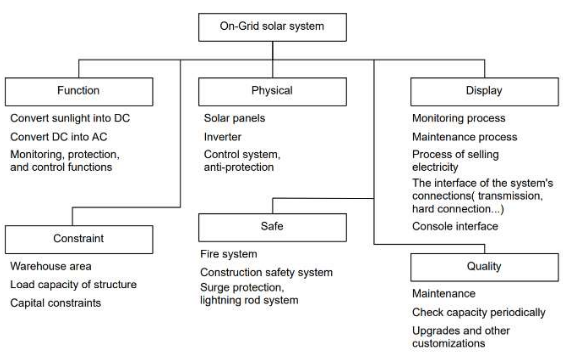


Figure 9. Functional baseline of the solar energy system

Table 9. System requirements specification table

<table>
<tr><th colspan="2">SYSTEM REQUIREMENTS SPECIFICATION TABLE</th></tr>
<tr><td colspan="2">Name</td></tr>
<tr><td colspan="2">Solar power system of ABC Company warehouse</td></tr>
<tr><td colspan="2">Description</td></tr>
<tr><td colspan="2">• The system is designed with the aim of being a new, friendly source of energy, strategically improving the cost of energy use for the company.
• The panel system is located in the attic with a large number of live loads in the bearing range, with a total capacity of over 2 MWP / year.
• The supporting frame system ensures durability and solid bonding
• Compact Transmission System has been marked with each installation area for easy inspection and repair in the future
• The inverter system is installed at the same station, the distance between the inverters, the electrical cabinet must ensure safety criteria, easy to disassemble and repair for warranty later.
• Layout of electrical circuits, electrical cabinets is simple, easy to control, manage ...
• The monitoring system and safety equipment operate automatically, must handle unexpected situations, disconnect when the power grid goes out.</td></tr>
<tr><td colspan="2">Function</td></tr>
<tr><td colspan="2">The system has 3 main functions:
• Converts photovoltaic energy into direct current
• Converts direct current to alternating current
• Control functions, connecting power source to the national grid
• In addition to the above main function, a number of subsystems undertake the overall protection and supervision function ... and systems with the function of transmitting throughout the main functions.</td></tr>
<tr><td colspan="2">Effective properties</td></tr>
<tr><td colspan="2">The system in operation must satisfy the following criteria:
• Ensuring technical requirements according to the circular of Ministries and Branches.
• Ensuring production capacity meets strategic goals, payback in 5-10 years
• Ensuring overall durability and safety.</td></tr>
<tr><td colspan="2">Attachments</td></tr>
<tr><td>• Drawing of warehouse layout, truss column structure
• Module capacity test table
• Job description sheet
• Work roster
• Table structure analysis factory</td><td>• System operation process
• Order and registration procedures
• Payment process
• System maintenance process and corresponding forms</td></tr>
</table>

6. Preliminary design

6.1. Allocation Matrix

Based on the system specification above, there are 11 functional requirements of solar panel system for warehouse. The system includes 03 groups of subsystems: physical, safety and interface system listed as below with attached numbers corresponding together with the content of the horizontal row of the below table.

- The physical system includes the functions of physical structures such as: Solar arrays (1), Transmission systems (cable networks) (2), Supporting frames and shelving (3), Electric meters (4), On-grid inverters (5), Monitoring modules (6), Switchboards to protection (7), Power distribution cabinets (8).

- Safety includes 03 items: Lightning protection system (9), CB anti-shock system (10).
- The interface includes the connection procedures between related departments: Registration process (11), Payment process (12), Maintenance process (13).

The main functions of the solar panel system for warehouses, which work together in accordance with the subsystems, are presented in the below Table 10.

Table 10. Allocation matrix

Allocation Matrix	Physical								Safety		Interface		
	1	2	3	4	5	6	7	8	9	10	11	12	13
1. The installation area is 24.000 m².	X	X	X	X	X	X	X	X	X	X			
2. Cost is less than 30 billion VND.	X	X	X	X	X	X	X	X	X	X			
3. Light can be received.	X												
4. Possible to covert energy.	X				X								
5. It can support, shape the system, the subsystem.		X	X										
6. The system can convert direct current (DC) to alternating current (AC).					X								
7. The system can delivery electricity.		X		X	X		X	X					
8. The system can distribute AC power.								X				X	
9. The system can measure AC power.				X								X	
10. The system ensures electrical safety.						X	X		X	X	X		X
11. The system can cut off power protection when the power grid stops supplying power							X	X			X		

6.2. Define Connection Interface

Interface can be divided into hard and soft interfaces. Hard interface includes control, monitoring and feedback mechanisms based on electronic hardware devices equipped together with the equipment packages of the system. While soft interface involves human interventions such as operational policy, maintenance, and specific defined duty contracts. Within the limit of researh, they are listed and shown as below:

6.2.1. Hard Interface

Monitering system: Ensuring a safe operation for a system is extremely necessary. For the solar power system, a monitoring system can be used with an additional RFID sensor installed in monitoring operation. This system will periodically evaluate each component of the system and centrally administer on a database / knowledge base. The monitoring system works continuously by sending periodic commands (for example, 5 minutes, 10 minutes, 60 minutes) to all sub-systems that need to monitor, receive results, store and analyze with history comparison, gives necessary warnings to the right people to handle the reaction. In some simple cases, the warning system can also fix the problem by itself according to a known scenario and report it to the people later.

Solar panel outdoor measuring system: Used to determine maximum capacity, measure temperature coefficient, check normal operating cell temperature NOCT (NMOT), test low irradiation efficiency, test outdoor exposure, test durability, test temperature.

Solar radiation sensor: One of the most important and popular solar power monitoring devices. They monitor the amount of solar radiation in the area where the panels are installed, helping the factory to optimize the absorbed energy of the panels as well as make necessary adjustments and assessments.

6.2.2. Soft Interface

There is a need to use a rooftop solar panel power sale contract between the user and the solar power factory. Between solar power factory and state management agencies, annual written reports are required. There are also rules during construction process.

6.3. Design Subsystems

As mentioned in the Table 11 - Allocation Matrix, the solar energy system consists totally of 14 subsystems. But in the limit of the research paper, we just mentioned the design process of 4 main subsystems that have the most significant impact role on the system. They include Solar arrays, On-grid Inverters, Lightning protection system and Maintenance process.

6.3.1. Solar Arrays

Below flow chart shows the basic steps to gradually calculate first subsystem - solar panel. This process starts with the basic element of solar panels, the selection of the number, then arranged them into string, many strings into arrays and final result of appropriate device performance layout as shown below Figure 10.

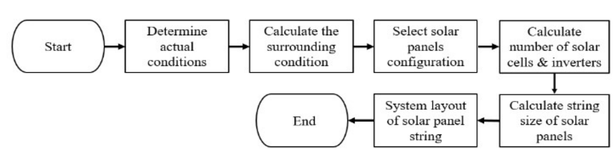


Figure 10. Calculating steps of the sub-system flowchart

Let x is the number of solar panels in string. This is Solar panels calculation (string size).

Limit of Solar cell chain voltage:

$$\frac{V_{mpp(T)\min}}{V_{mp\_solar\_panel}} < x < \frac{V_{DC\max}}{V_{OC\_solar\_panel}} \quad (1)$$

Optimal range for rated capacity:

$$V_{mpp(T)\min} < x \times V_{mp\_solar\_panel} < V_{mpp(T)\max} \quad (2)$$

Normal working temperature limit:

$$\left(V_{OC} + \left[V_{OC} \times H_{VOC} \times \left(T_{\max} - T_{\min}\right)\right]\right) \times x < V_{DC\max} \quad (3)$$

Constraint (1) usually used to estimate number of solar panels, while constraint (2) and (3) has role of more accurate assessment for the results of (1). Converting all variables to right side, we have a set of constraints for the number of solar panels as follows

(1), (2) and (3) <=>

$$\begin{cases} \dfrac{V_{mpp(T)\min}}{V_{mp\_solar\_panel}} < x < \dfrac{V_{DC\max}}{V_{OC\_solar\_panel}} \\ \dfrac{V_{mpp(T)\min}}{V_{mp\_solar\_panel}} < x < \dfrac{V_{mpp(T)\max}}{V_{mpp\_solar\_panel}} \\ x < \dfrac{V_{DC\max}}{\left(V_{OC} + \left[V_{OC} \times H_{VOC} \times \left(T_{\max} - T_{\min}\right)\right]\right)} \end{cases}$$

For large-scale projects, it is neccesary to consider using the maximum number of solar panels for the chain that satisfies the above constraint system. For example, with 2.5 MWP system need using 6,250 units of solar panel with 400 W wattage. If We using string 15 panel, the number of strings in system reach the number 417. In the choice of the largest capacity inverter available today, need 23 inverters for all. Meanwhile, just use 17 inverters with string 20 panels.

The reason there is no limit on the amperage is in the series connection of solar panel, there is no change in amperage. The Amperage of string always equal to the panel. Almost solar Inverter support for this style of connection.

The system will select solar cells 400W and Inverter 110kW to perform the calculations below. The specifications of the 2 devices are shown in below tables.

Table 11. Specification of Inverter 110kW

Input specification (DC)	
Maximum input voltage	1100V
Minimum input voltage / Start	200 V/250V
Input voltage level	585 V
MPP voltage range	200V–1000V
MPP voltage range for rated capacity	550V–850V
Number of independent MPP inputs	9
Maximum number of strings per MPPT	2
Maximum current per input	30A
Maximum input current	26A*9
Maximum current for input connection	30A
Maximum DC short-circuit current	40A*9

Table 12. Specification of Inverter 400W

Minimum performance at standard testing conditions		**Minimal performance under normal operating conditions**	
Capacity of MPP	400	Capacity of MPP	299.4
Short-circuit current (Isc)	10.19	Short circuit current Isc	8.21
Open Circuit Voltage (Voc)	48.96	Open circuit voltage Voc	46.17
Current in MPP (Imp)	9.7	Current in MPP	7.64
Voltage in MPP (Vmp)	41.23	Voltage in MPP	39.20
Conversion performance	≥ 19.9		
Temperature coefficient			
Thermal coefficient of Isc [%/K]	+ 0.04	The Heat Coefficient of Voc [%/K]	-0,27
Thermal coefficient of Pmp [%/K]	- 0,36	The module's operating temperature is normal (ºC)	43±3
Specification of each unit: • Measurement tolerance PMPP ± 3%; Isc; • Voc ± 5% ở STC: 1000W / m2, 25 ± 2°C, • AM 1.5G According to the standard IEC 60904-3 • 2800W / m², NMOT, quang phổ AM 1.5G			

About the structure of each unit of solar panel, it includes multi-layers typically shown as below Figure 11.

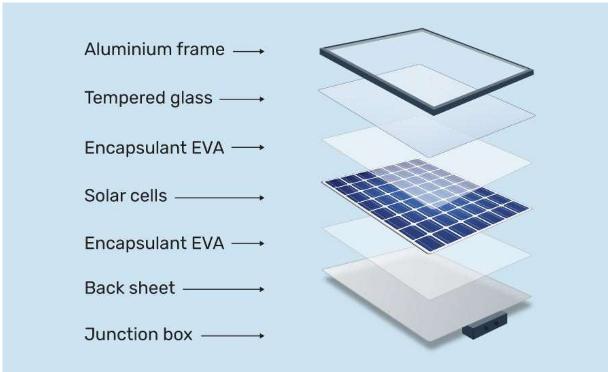


Figure 11. Structure of Solar array (SolarReview, 2021).

The solar arrays include many solar panels arranged together, connected by power cable, play the role of collecting and converting solar energy into electricity. They are the main components of the solar panel system. Solar arrays are divided into 8 main parts including: Aluminum frame, tempered glass, encapsulant - EVA, solar cells, back sheets, junction box, power cable, MC4 connector.

The solar arrays will be permanently installed on the roof through the racking frame system (Aluminum Rail/Steel Frame Truss). The solar arrays will be connected to each other through the built-in MC4 ports on the back of the panels.

There are 3 methods of pairing solar cells:

- Serial connection: increase the total voltage of the chain.
- Parallel connection: increases the current intensity of the chain.
- Combination of parallel and serial connection: increase both voltage and current intensity of the chain.

All three methods increase the total system capacity, but with the parallel connection method, current intensity can reach a very large value, exceeding the rated value of many Inverter systems while nowadays many manufacturers develop Inverters with high operating voltage to support serial connection systems.

The material of the solar arrays has 4 options: mono (monocrystalline), poly (polycrystalline), thin-film and mono Perc (improved version of the mono panel).

The table below shows a comparative evaluation of options based on capacity and cost.

Table 13. Comparison between material selection options

Items	Mono	Poly	Thin-film	Mono Perc
Defina-tion	Monocrysta-lline solar panels contain cells that are cut from a single crystalline silicon ingot.	Polycrystal line solar cells are made from ingots of molten, quenched, and hardened silicon.	Thin film solar cells are made up of very thin films of molten silicon material.	Improv-ed version of the mono panels.
Capacity	Highest perform-ance (24.4%)	High perform-ance (19.9%)	Average performan ce (18.6%)	High perform ance (greater than 20%)
Cost	Expensive 36 USD / 60W plate	Afford-able 31 USD / 60W plate	Cheapest 13 USD / 60W plate	Afford-able 31 USD / 60W plate
Charac-teristics	Fast sunlight absorption, as long as it is light this panel generates electricity	Absorb-ing sunlight quite slowly. It stops working when the weather is cloudy.	Lighter weight and more maneuve-rable, making it easier for installers to carry panels on roof.	Improv-ed version of the mono panels.

After comparing options, the team evaluated the options on a 10-point scale with 3 criteria of cost, capacity and efficiency (with weights of 0.3, 0.3, 0.4, respectively).

Table 14. Trade-off for Solar arrays material options

Criteria	Weights	Mono	Poly	Thin-film	Mono Perc
Cost	0.3	5	6.5	8	6.5
Capacity	0.3	8	6	5.5	8.5
Efficiency	0.4	7	5	6	8
Total	1	6.7	5.75	6.45	7.7

After evaluation, Momo Perc was the material selected with the highest score (7.7). With 2 types of equipment has been selected, we proceed to calculate the number of chains for the solar array subsystem.

Determine the optimal battery string size:

$$\frac{V_{mpp(T)\min}}{V_{mp\_solar\_panel}} < x < \frac{V_{DC\max}}{V_{OC\_solar\_panel}} \quad (1)$$

$$\rightarrow \frac{550}{41.23} < \mathrm{x} < \frac{1100}{48.96}$$

$$\rightarrow \quad 14 \leq \mathrm{x} \leq 22 \ (1)$$

Optimal range for rated capacity:

$$\frac{V_{mpp(T)\min}}{V_{mp\_solar\_panel}} < x < \frac{V_{mpp(T)\max}}{V_{mpp\_solar\_panel}}$$

$$\rightarrow \frac{550}{41.23} < x < \frac{850}{41.23}$$

$$\rightarrow 14 \leq x \leq 20 \ (2)$$

Calculate the temperature of the battery system:

- The battery module's normal operating temperature (NOCT) is 43 ± 3 C^o.
- The lowest temperature can be reached in Nam Tan Uyen Industrial Park: 18 C^o.
- The maximum temperature difference: 28 C^o.
- Coefficient of Voc: 0.27%

$$x < \frac{V_{DC\max}}{\left(V_{OC} + \left[V_{OC} \times H_{VOC} \times \left(T_{\max} - T_{\min}\right)\right]\right)}$$

$$\Leftrightarrow x < \frac{1{,}100}{\left(48.96 + \left[48.96 \times 0.27 \times \left(28\right)\right]\right)}$$

$$\rightarrow x \leq 20 \ (3)$$

Consider system characteristics: With xmax and results (1), (2), (3), we have x = 20.

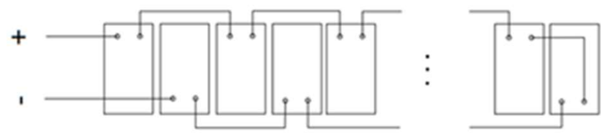
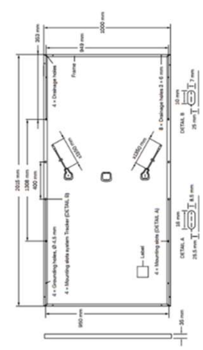


Figure 12. The solar array connection layout

Table 15. Specification of solar panel in calculation

	Size:	2015mm x 1,000mm x 35mm (including frames)
	Mass:	23kg
	Junction box:	32-60mm x 53 - 101mm x 15-18mm
	Cable:	4mm^2, (+) ≥ 1350mm, (-) ≥ 1,350mm
Total number of solar panels:		5,760 units
Total installed charge of solar panels:		2,015x1x 5,760 = 11,606.4 $\boldsymbol{m^2}$
Total capacity of the system:		400×5,760=2,304,000 Wp≈2.3 mWp

6.3.2. On-grid Inverters

6.3.3.

From the DC electrical cabinet, the strings will be connected to the inverter. Inverter will have maximum capacity optimizer called MPPT. Each MPPT will include 1 or more +/- input pairs. With 1 MPPT Inverter, only 1 solar array can be installed. In addition, the panels will need to be installed on the same plane and inclination. With 2 MPPT Inverter you will be able to install 2 or more solar arrays in two different directions. The number of MPPT / DC inputs also affects the conversion efficiency of the inverter. To convert the direct current (DC) collected by the solar system, convert it into alternating current (AC), using on-grid Inverter. When not being confident in the monitoring system, specialized equipment, a single-phase power meter, will help verify this. The device is mounted on the front of the AC switchboards to protection.

The arrangement of the inverter must both ensure the requirement for easy system management while minimizing costs. Each 20 panels string includes 2 DC + and DC- wires, 110kW inverter supports 18 strings (9 pairs of DC input 2 in 1). With the above conditions, the specification of inverter has been chosen and shown as Table 17.

Structure diagram shown the relationship of the PV array - Inverter - Transformer - Electrical Grib is below Figure 13. The arrangement of Solar Panel Arrays on the roof of warehouses and the location of the Inverter Station are illustrated in Figure 14.

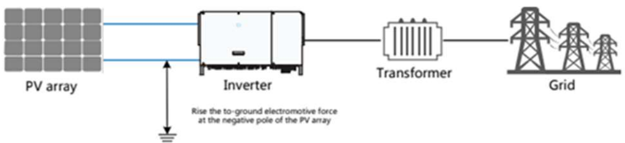


Figure 13. The solar array connection layout

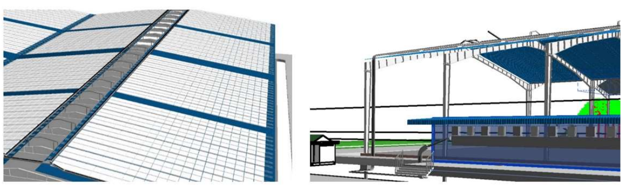
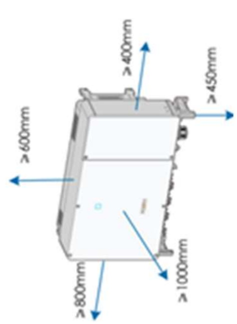


Figure 14. 3D illustration photos of solar panel arrays (on the building roof)

Table 16. Specification of Inverter



Technical specifications	
Dimensions:	1,051 × 660 x 362.5 mm
Mass:	85 kg
Operatingtemperature range:	-30 –> 60 °C
Standard AC connection:	OT terminal
Efficiency:	98,7 %
Amount:	12 unit

6.3.3. Lightning Protection System

Before designing a lightning protection system, it is necessary to clarify the feasibility of lightning protection system, particularly in the U&I warehouse area.

Table 17. Feasibility analysis of lightning protection system

No.	Factor	Notes
1	***Frequency***	The warehouse location has been recorded with high risk of annual rainfall and lightning hazard.
2	***Availability***	The system is not equipped with lightning rods or any other lightning protection systems.
3	***Scale of construction***	Large.
4	***Kind of system***	Grid-tie system.
5	***Location***	Located in industrial zone.
6	***Rated voltage***	Very high operating voltage.

In addition to the surge protection system designed in electrical cabinets, the design team proposes a direct lightning protection solution for the solar energy system.

Install lightning rod according to early emission technology, mounted on external independent pillar. These active air-termination needles have a very large protection radius (from 50-107 m), the number of lightning rods is arranged so that its protective zone covers the entire surface of the system. For these large buildings, if lightning dissipation needles are used, the better it will disperse the opposite charges to prevent lightning currents from hitting the area, which is safer but the cost may be higher. Using charge dispersion technology will prevent lightning from hitting the area it protects, this effect means that it will limit the occurrence of repeated overvoltage pulses or lightning propagating on the line of solar panels system. Construction and installation of the grounding system for these lightning rods must meet the technical criteria, the maximum ground resistance value is 8 Ohm. Grounding systems also need to be equipotentially linked.

Regarding surge protection solutions for DC solar power and for AC power lines: According to CLS / TS 5039-12 recommendation, use DC power surge protective devices to protect on lines connecting from panels to the power cabinet, lightning protection devices for DC solar power must be installed for protection at the DC entrance of Invertor and the panels. The CLS / TS 5039-12 standard also recommends the installation of AC mains surge protective devices to protect directly in front of the Invertor's AC entrance, of loads and breakers connected to the mains. Overvoltage can spread on the line from loads (pumps, lights, etc) or from the national grid (connected to the network).

The grounding system is an extremely important part of the system, it has the function of working earthing, protective earthing and lightning protection earthing. All structures of supports, cabinets, frames, lightning protection devices must be grounded, ensuring equipotentialization in the whole system, equipotential devices (such as AT-50K) can be used to link grounding zones for different functions. The smaller the ground resistance value, the better (maximum 8 Ohm) depends on geographical features and capacity of the solar panels system. The materials used in the solar grounding system should be of a good type, have high corrosion resistance (because it is affected by chemical corrosion, electrolytic corrosion).

The ground piles can be of high-quality copper or pure copper, especially if chemical ground piles (ApliRod) are used, the Graphite ground piles will ensure the highest efficiency. You can use more chemicals which can reduce the soil resistance to enhance electrical conductivity, reduce resistance such as Conductiver Plus (used for sandy, rocky soil), Aplifill (used for electrode pits), Aplicem (used for piles and bonding wires). The number of earth electrodes and chemicals used depends on the specific geological features of each project. Links between piles, wires and other metal structures should use thermochemical welds (such as Apliweld) to ensure the best electrical bonding, as well as to ensure mechanical properties and durability in corrosive environments.

The table below shows the basic technical requirements of a direct lightning protection subsystem for solar panels system.

Table 18. Technical requirements of components in lightning protection system

No.	Components	Technical requirements
1	Lightning rod	- Vietnam's lightning protection design standard: TCXD 46: 2007. TCVN 9385-2012 standard.
2	Collectors and bases	- Made of high strength stainless steel, - Has high strength, meets the standards for building diffusion systems.
3	Bonding wires	- The down-conductor directly from the lightning protection needle or the ground cable from the main earthed copper plate will be linked to the earth system at the central pile location (earth resistance test hole location). - Large load current, small impedance.
4	Ground well	- Applicable to soils with high soil resistivity - Good diameter is defined in range of from 50mm to 80mm. - The depth is from 15-40m depending on the depth of the groundwater
5	Ground rod	- In the case of drilling a well, the ground rod will be directly linked with the cable to drop deep into the bottom of the well. Pouring chemicals can reduce soil resistance to the well, and pouring water down so all chemicals can settle down to the bottom of the well.
6	Others	- Insulating plastic base, resistance reducing powder, wire harness, anchors, thermochemical solder, explosives etc.

6.3.4. Maintenance Process

Solar power system has been recommended an adequate periodical maintenance to ensure that there has been not any problems in the operation circle time of the system followed both popular maintenance policies of protection and prevention based on the system equipments, conditions and capabilities.

The policy has been implemented and simplified as below Table 9, listed items can be conducted by in-house technicians of the owner, combining together with the observation by remote devices and maintained by outside constractors.

Table 19. Solar energy system maintenance table

Object	Forms of maintenance	PIC department
Solar panels	- Checking for rust, damage observed on the face of solar panels. - The glass on the panel may break - Testing against water permeability. - Testing capacity	Visually checked by in-house technicians
PV system (MC4 connector)	The connection is still good, not leaking	Outsourced
Solar battery system	Sweeping and cleaning the battery	Internal technicians
Grounding & Anti-lightning	Grounding and anti-lightning testing is conducted periodically.	Outsourced
Damaged roof	Replacing the damage metal roof in case of there is any damage or rust phenomenon	Outsourced

	happened due to weather effects	
Vent	Cleaning	Outsourced
Electric circuit breaker system	Checking for normal operation	Outsourced

7. Detailed design

7.1. Solar Panel

The panels play the role of collecting and converting solar energy into electricity. They are the main components of the solar cell system. There are 3 methods of solar cell coupling (serial, parallel, mixed), all of 03 methods increase the total system capacity.

However, after many calculations, considered the requirement of owner and expert opinion, the structure of 18 parallel chains, each consisting of 20 batteries has been chosen for the new system with the specification of solar panels followed the commercial products under different brand names shown as below Table 20.

Table 20. Specification of solar panel under brand name in the marketplace

No.	Product Name	Technical Specification	Price
1	Brand A	- Wattage: 400Wp/unit - Type: Monocrystalline (72cells). - Module Efficency: 19.3% - Size: 2024 x 1024 x 40 (mm) - Weight: 21.7 kg - Product warranty period: 15-year product warranty and 25-year linear performance warranty	390 USD per unit
2	Brand B	- Wattage: 400Wp/unit - Type: Mono – Half cut cells (144cells). - Module Efficency: >=19.9% - Size: 2015×1000×35 (mm) - Weight: 23.5kg - Product warranty period: 12-year product warranty and 25-year linear performance warranty	208.24 USD per unit
3	Brand C	- Wattage: 400Wp/unit - Type: Monocrystalline (72cells). - Module Efficency: 19.88% - Size: 1979x1002x40 (mm) - Weight: 22.5 kg - Product warranty period: 15-year product warranty and 25-year linear performance warranty	158.26 USD per unit

Brand B products has been recommended for the owner and stakeholders based on the advantages of durability, warranty policy, light weight compared to battery models with the same capacity and efficiency.

7.2. Inverter

The use of high-power inverters is more efficient at the expense of a lower DC input, less installation space and easier system control.

Table 21. Specification of inverter under brand name in the marketplace

No.	Product Name	Technical Specification	Price
1	Brand X	- Maximum wattage: 12000W - Conversion efficiency: 98.4% - Size (W / H / D) mm: 661 x 682 x 264 - Mass: 61 kg	3,000 USD per unit
2	Brand Y	- Maximum wattage: 110kW - Conversion efficiency: 98.7% - Size (W / H / D) mm: 1,051 × 660 x 362.5 mm - Mass: 85 kg	5,046 USD per unit
3	Brand Z	- Maximum wattage: 55kW - Conversion efficiency: 98.7% - Size (W / H / D) mm: 782 × 645 × 310 mm - Mass: 62 kg	2,814 USD per unit

Brand Z products has been recommended for the owner and stakeholders based on the advantages of low cost, high conversion efficiency, and proven stability.

Through the calculation of the number of solar panels and the unit price as above, we have a preliminary estimated purchasing of 1.19 mil USD for solar panels (5,760 units - 208.24 USD per unit) and 0.045 mil USD for inverters (12 units - 2,814 USD per unit) respectively. Thus supporting equipments, installation labor cost, and others, final quotation can basically be lower than the investors expected limit (about 1.3 mil USD).

Illustrating the design results, we would like to show the detailed electrical drawings of the solar energy system, shown as below Figure 15.

7.3. Operation and Maintenance (O&M) Procedures

When beigning operation, the new system need to be monitored, and then be periodically maintained to ensure the system is operated continuously, minimizing possible hazards, and increase the life expectancy of machinery systems.

Expected results have been defined as managing owner's assets effectively, optimize operating costs for the main occupancy of the owner's premises, even in the post-warranty period. In the limit of the research paper, the main items has been mentioned and simplified to make the basis for implementation steps when the designing has been developed actually. The operation and maintenance process includes many items shown as below Table 22:

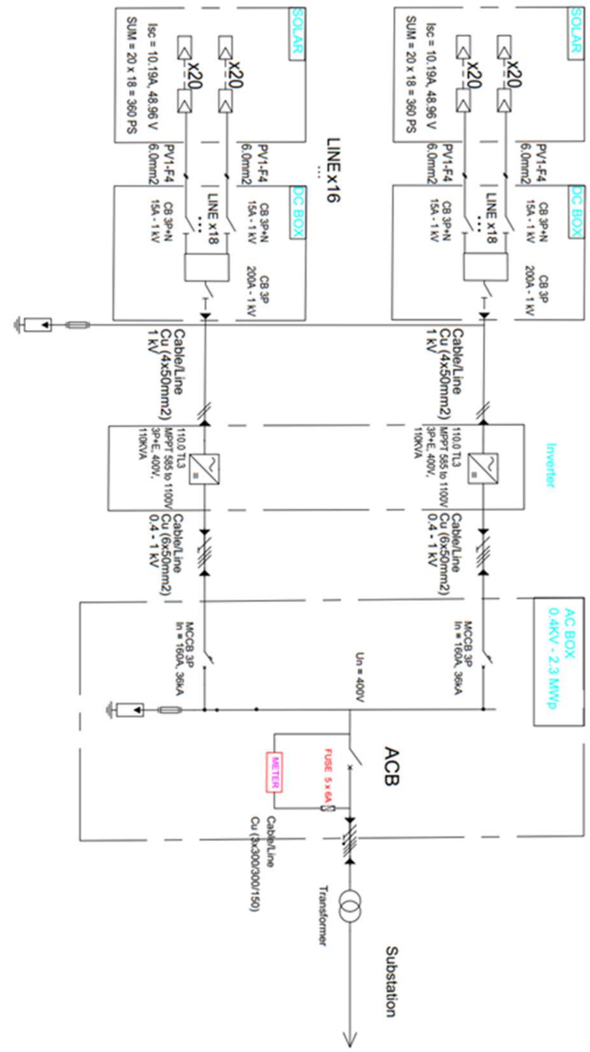


Figure 15. Single-line diagram of the new solar energy system

Table 22. Operation and maintenance process of the new solar energy system

Items	Detailed implementation
Supervision and monitoring	- Monitoring the performance of the solar power system. - Error detection and system diagnostics. - Periodic inspection and assessment at the warehouse.
Spart part storage	- Production system recovery. - Repair and replacement of equipment. - Manage spare parts.
Preventive maintenance	- Maintain Inverter, connection points, filters, fuses, upgrade and update hardware, prepare for commissioning and commissioning. - Maintenance of DC connection cabinets, connection points and watertight test. - Tracker system: Check and lubricate. - Battery plate: Check the battery plate and connection points. - Check the frame system to support the panels. - Circuit: Monitor, check thermal image. - Check transformers and switches. - Surveillance and security system.
Grid integration adjustment	- Control power, voltage, reactive power and frequency. - Forecast of electricity output. - Schedule a power cut-off for maintenance purposes. - Identify system trends and report analysis. - Make forecasts. - Plan and take action to prevent.
In-plant maintenance	- Cleaning solar panels - Ensuring factory security.
Technical services	- Prepare for trial run, test run and re-run. - Quality control and system inspection. - Upgrade and improve the system. - Set up monitoring system. - System improvements.
Property management	- Manage the output of the battery system. - Manage the performance of the battery system.

8. Conclusion & Recommendation

During the study and design process, the research has been prepared for many scenarios and situations as possible to adopt the construction, operation conditions and maintenance of the system. So the output product of the rooftop solar PV system design process has met the requirements given by all stakeholders, proven feasibility, highly applicable and suitable with the actual conditions of the owner premise. However, the reality will always have complicated developments, outside of the group's expectations. Then, the initial plans can be flexible to best suit the external conditions, the actual situation as well as the changes required by stakeholders, if any.

In case of the rooftop solar PV system is applied to the warehouse system and operates well in theory, achieves the expected high efficiency, saves on electricity costs and can be a source of profit from residual electricity Uploaded to the State, in the future, the design team proposes to apply more advanced science and technology to invest in an intelligent management system for the system. Generally, the rooftop solar PV system will bring many economic and environmental benefits, the system also guarantees safety, usage time and payback time to adopt the owner requirements.

However, due to the limited level of research time and other objective conditions, the group has not been able to apply many science and technology to the system in the actual machinery installment process of new system.

Reference

Anh, N. D., Linh, Y. T., & Mai-Ha, P. (2020). Warehouse Design For Electrical Equipment Manufacturer: A Case Study. *International Journal of Mechanical Engineering & Technology (IJMET) 11*(7), 51-71.

Arno, S., Klaus, J., Olindo, I., Rene, V. S., Miro, Z. (2015) - *Solar Energy - The Physics and Engineering of Photovoltaic Conversion Technologies and Systems*, England: UIT Cambridge

Dutoan.cic.com. (2021). Software for calculating and designing pre-engineered steel frames (Zamil) - KIW. Retrieved from https://dutoan.cic.com.vn/phan-mem-tinh-toan-thiet-ke-khung-thep-tien-che.html

Kossiakoff, A., Sweet, W. N., Seymour, S. J., & Biemer, S. M. (2011) - *Systems Engineering Principles and Practice*. New Jersey, USA: John Wiley & Sons.

Mitsubishi pulls out of Vietnam coal plant amid climate concerns. (2021). Retrieved from https://asia.nikkei.com/Business/Energy/Mitsubishi-pulls-out-of-Vietnam-coal-plant-amid-climate-concerns.

Oliver, D. W., Timothy, P. K., & James G. K. J. (1997) - *Engineering Complex Systems with Models and Objects*. New York, USA: McGraw-Hill.

Philipps P. S., Bett W. A., Horowitz K., & Kurtz S. (2017) - Current Status of Concentrator Photovoltaic (CPV).

Technology, Technical Report Version 1.3, Fraunhofer Institute for Solar Energy Systems ISE.

SolarReviews.com. (2021). How do you build your own solar panel system?. Retrieved from: https://www.solarreviews.com/blog/how-to-build-your-own-solar-panel-system

Thang, T. N., & San, T. T. (2014) - *Solar Energy Handbook*. Hanoi, Vietnam: Science and Technics Publishing House.

AN EMPTY CONTAINER OPTIMIZATION IN DISTRIBUTION NETWORK OF A LOGISTICS COMPANY UNDER UNCERTAINTY - THE CASE OF ASIA

Hoang Thi Bich Ngoc[1]

[1]*Quy Nhon University, hoangthibichngoc@qnu.edu.vn*

Abstract: There is an unpredictable shortage of shipping containers in Asia, driving up shipping costs and delays for goods purchased from Asia to other parts of the world. The economic rebound from the corona virus pandemic is being blunted by a global shortage of shipping containers, sending cargo costs to record highs and hampering manufacturers in filling fast-recovering global goods orders. Delays and contributing to higher freight rates from Asia to Europe and the USA are largely caused by a shortage of shipping containers in major export ports. The shortage is further exacerbated by limited air freight capacity. An urgent question asked is how container carriers could minimize container costs, particularly container repositioning costs and innovative solutions as well as optimization techniques to improve the efficiency and productivity of yard operations could be taken into consideration. The study will find and analyze reasons to the shortage of export containers in Asia and forecast demands and supply for containers. The paper also paper provides an optimization method for the intermodal transportation of full and empty containers in harbor-inland regions/hinterlands. The model could be proofed to be feasible in the real transportation networks and can create more profits for transportation enterprises. In particular, the study deals with analyzing the distribution of containers in the yard layout as a function of the moment at which space for export containers is reserved while looking at the operational costs; optimizing the dimensions of a perpendicular layout and addressing the problem of empty container repositioning (ECR) in the distribution network of logistics companies. The study also emphasizes how the opportunities in technological development can be used in ECR practices and provides suggestions for research topics that can be studied in the light of current trends in container shipping.

Keywords: Container logistics; Container shortage; Storage capacity; Empty container repositioning.

1. Introduction

Approximately 90% of world trade are carried out using sea transportation and shipped using containers (Bandeira, 2009). Asia also remained the leading global player in container port cargo handling, with a share of nearly 65% of global container port traffic (UNCTAB, 2020). Because the most Asia nations in Asian remain highly dependent on oversea markets as well as closely connect to the global economy through global value chains, shipping represents as a lifeline for social and economic development in Asia. The most major container ports are in Asia and the large number of these ports is located in China. Therefore, the transport infrastructure and logistics corridors associated with the intermediate services (highways, railways, barges and short sea shipping) have developed for these ports to participate in the polarization of maritime traffic. Shipping lines here fill containers from inventory in the form of empty containers owned by shipping lines at the port. Shipping empty containers to other ports is known as repositioning, an important phenomenon in the container shipping industry. Empty container repositioning (ECR) aims to pick up unused empty containers from ports where they are in surplus and providing them in ports where they are in shortage. Many researchers developed a repositioning model through considering transportation topology. Du and Hall (1997) developed the single threshold policy model using a hub-and-spoke system was developed by Du and Hall (1997), while Li and Liu K (2004) developed the two-threshold policy for a single port, then expanded to a two-threshold policy for a multi-port system. The combination of cargo routing model with ECR multi services routes, multiple deployed vessels and multiple regular voyages was developed by Song and Dong (2012). Technological advances have succeeded in creating containers that can be folded so that repositioning does not require a large place. A mathematical model to minimize the cost of repositioning both foldable and unfoldable containers were developed by Myung Y-S (2014). Wang K (2017) made numerical experiments to compare compare the total cost between foldable and unfoldable. ECR has been an on-going issue since the beginning of containerization and will remain as a key issue in the future due to the nature of the industry.

What makes empty container management particularly cumbersome is data uncertainty. However, decisions on ECR are taken daily, even if data are affected by uncertainty. As an example, the uncertainty may concern the empty container demand arising from unexpected transportation opportunities. Similarly, this paper focuses on demand uncertainty in the period of COVID 19 pandemic. The uncertain number of empty containers requested in ports, because transportation requests may unexpectedly pop up, be postponed and cancelled. These events can change the optimal volume of empty containers to be repositioned between ports. From the logistics channel scope perspective, the ECR problems could be tackled internally within the shipping company, externally in the vertical logistics channel, externally in the horizontal channel (i.e. collaboration with other shipping companies), and through technological innovations (Mehmet Yasin Göçen, 2020). Those joint optimization problems are often complicated. Most of them use heuristics/meta-heuristics to tackle the problems or model the empty container repositioning in less detail to make it analytically tractable.

The below fishbone diagram (see Fig. 1) shows the issues of the current system, resulting in the necessity of repositioning empty containers to satisfy the inventory constraints for empty containers. Furthermore, minimizing empty container movements will reduce external effects of transport such as congestion and air pollution, which makes the topic even more relevant.

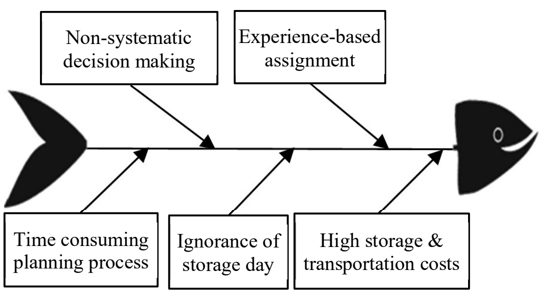


Figure 1. Fishbone diagram (Göçen, 2020)

Underestimating empty container demand may contribute to the accumulation of unnecessary empty containers in import-dominant ports, because they may not be shipped in a timely fashion where they are requested. On the other hand, demand overestimation results in unnecessary repositioning costs. In order to minimize both operational costs and shortage costs while meeting customer demands for empty containers, logistics companies must decide how many empty containers to load/unload to/from vessels, and how many empty containers transport in the residual transportation capacity.

The aim of paper is (i) Identify managerial systems associated to container repositioning flows; (ii) find and analyze reasons for the shortage of containers in Asia and (iii) propose a solution to the situation, especially in focusing on modeling of empty container repositioning. More specifically, the ECR problems will be able to tackled internally within an organization, externally in the vertical logistics channel, externally in the horizontal channels (i.e. collaboration with other shipping companies), and through technological innovations.

2. Problem description

In the existing system of Asia, full containers could reach a consignee's premises directly by truck through a marine container terminal or multimodal transport. Once stripped, empty containers are either returned to the marine terminal or a storage depot, or shipped directly to a consignor's premises for packing or backhauling load. At that time, containers become empty and needs to be cleaned and then, be transported to one of the available warehouses by the liner shipping company's trucks and stored there for a while. Occasionally, empty containers could be interregional repositioned through a depot or an intermodal terminal. They also reach consignors' premises through intermodal transportation and last mile drayage. Once filled at the consignor's premises, containers are drayed to an intermodal terminal or a marine terminal for export. Before overseas repositioning takes place, empty containers hired by an ocean carrier reach storage depot and be temporarily stored. From storage depots, aged containers, particularly those stored over a long period, are sold out of the transportation network to the secondary market. Empty containers may move between marine terminals and storage depots or between different storage depots for balancing purposes. Some terminals operate satellite empty container depots to gain additional storage capacity in order to avoid the congestion at the gates and provide dedicated service to ocean carriers.

The below figure (see Fig. 2) represents the dynamics and interactions among stakeholders at the local and regional level. The demand for integration between global and local container operations is clear, considering the unbalance proportion.

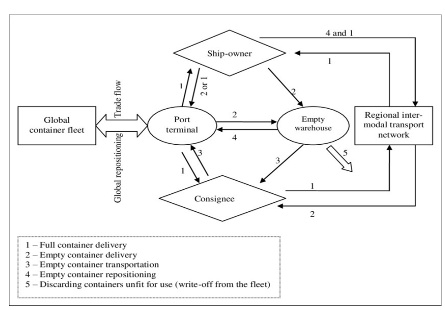


Figure 2. Container flow and repositioning (Theofanis & Boile, 2009).

The global transportation system is currently experiencing a unique and unexpected crisis when it faces a serious problem: the shortage of empty containers, especially in Asia. The lack of containers in the right locations during the recovery phase had a massive impact on rates. The container shortage also drives up new containers' prices, as manufacturers know the demand to charge extra. Asia container manufacturers, which dominate the market, are now charging $2,500 for a new container, up from $1,600 last year. This also has led to the cost of shipping a container globally to record highs. Freight carriers are facing a shortage of shipping containers amid a wave of demand for delivery by sea, helping drive up rates and increasing supply chain costs for some businesses across Asia. Shipping costs was initially less than 1,000 USD/40ft container has now increased to 8,000- 10,000 USD/40ft container (UNCTAB, 2020).

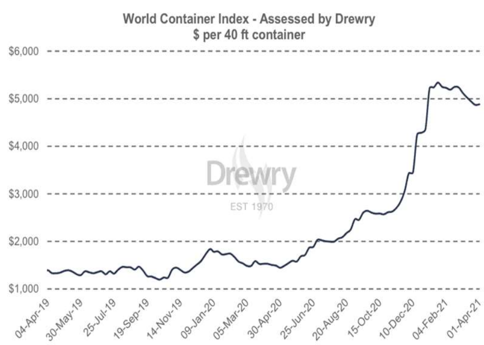


Figure 3. Word Container Index – $ per 40 ft. container (Assessed by Drewry, 2021).

The container crisis is also a reflection of a slowdown in the maritime supply chain due to strains caused by COVID 19 pandemic. The critical reasons that cause the crisis in Asia include:

Trade imbalance: This is currently the most important reason for the accumulation of empty containers in Asia. The reason means that a region that imports more than it exports will face the systematic accumulation of empty containers, while a region that exports more than it imports will face a shortage of containers. If this situation endures, a repositioning of large amounts of containers will be required between the two trade partners, involving higher transportation costs and tying up existing distribution capacities. Especially, changes in consumption and shopping patterns triggered by the pandemic, including a surge in electronic commerce, as well as lockdown measures, have in fact led to increased import demand for manufactured consumer goods, a large part of which is moved in shipping containers. Containers from Asia were sent to other nations, but due to COVID-19 restrictions, almost nothing moved in the opposite direction. Empty boxes were left in places where they were not needed, and repositioning had not been planned for.

Table 1. Containerized trade on major East–West trade routes, 2014–2020
(Million TEUs and annual percentage change)
(UNCTAD, 2020b)

Year	Trans-Pacific		Asia–Europe		Transatlantic	
	East Asia – N.A.	N.A. – East Asia	N.E. and Mediterranean to East Asia	East Asia– N.E. and Mediterranean	N.E. – and Mediterranean	N.E. and Mediterranean – N.A.
2014	16.2	7.0	6.3	15.5	2.8	3.9
2015	17.4	6.9	6.4	15.0	2.7	4.1
2016	18.2	7.3	6.8	15.3	2.7	4.3
2017	19.4	7.3	7.1	16.4	3.0	4.6
2018	20.8	7.4	7.0	17.3	3.1	4.9
2019	20.0	6.8	7.2	17.5	2.9	4.9
2020	18.1	7.0	6.9	16.1	2.8	4.7
Annual percentage change						
2014–2015	7.9	-2.0	1.4	-2.6	-2.4	5.6
2015–2016	4.4	6.6	6.3	2.5	0.4	2.9
2016–2017	6.7	-0.5	4.1	6.9	7.9	8.3
2017–2018	7.0	0.9	-1.3	5.7	5.8	6.8
2018–2019	-3.8	-7.4	2.9	1.4	-5.0	-0.2
2019–2020	-9.7	2.6	-3.6	-8.3	-5.3	-5.8

Table 1. shows that container flows on three major trade routes: Europe-Asia, Trans-Pacific (North America-Asia), and Trans-Atlantic (Europe-North America) routes. The trade demands in the Europe-Asia and the Trans-Pacific routes were severely imbalanced. In the Europe Asia and Trans-Pacific trade routes, European ports and American ports have been experiencing a high surplus of empty containers, while Asian ports are facing severe shortages. Despite the backlog, there are no signs of reduced trade between Asia and the rest of the world. A significant number of containers have found themselves in inland depots while others have been stacking up at cargo ports. As Asia slowly began to recover, other countries were still faced with national lockdown restrictions meaning containers could not be sent back to Asia, where they were needed to continue the trade partnership. There is currently a massive 40 percent imbalance in North America. This means that only four were sent back for every ten containers arriving, while six remained at the arrival ports. With trade between China and the United States averaging 900,000 TEU (20-foot equivalent unit) per month, the absolute container imbalance is genuinely enormous. The percentage of empty container movements in inland networks could be higher since empty containers are often

stored at ports or depots, which are away from the demand locations.

Port closures: Due to quarantine periods in effect, many ports have been closed in order to ensure the well-being of workers and various conditions have been imposed. Many countries also decided to impose a ban on vessel to be on the water or restrict the vessels' entry, leading to enormous disorganization between many shipping companies worldwide (Firm, 2020). As a result, the delay of goods transportation during quarantine led to a low capacity in ports, and the storage spaces were significantly crowded.

Lack of labor: There was a catastrophic labor shortage in American ports. And it was not just about docks and warehouses. Due to border restrictions, the work of customs was also partially suspended. Although China resumed exports earlier than the rest of the world, other countries faced and continue to face restrictions and labor cuts.

Reduced cargo: With reduced manufacturing, increased handling times, and large numbers of containers being held at ports, exporters reduced cargo. Deploying vessels without adequate cargo would have led to significant losses, putting shipping companies at serious risk. Therefore, logistics companies chose to reduce the number of ships in transit to retain some level of profitability. This slowed the flow of imports and exports and meant that only the most profitable trade routes were being used. The most significant example of this can be seen in the American regions, where Asian containers could not be sent back due to Covid-19 restrictions.

Finally, the situation was further exacerbated by the obstruction of the Suez Canal by a grounded container ship. As ships, and the containers on them, took longer to reach their destinations, the shortage of available empty boxes increased, including in Shanghai, China. As a result, freight rates increased not only on the routes passing through the Suez Canal but also on nearly all other routes.

3. Mathematical Model and Solution Methodology

The purpose of the part is to introduce a formulation that ensure an effective ECR for each different possible future evolutions. A multi-scenario formulation could be more appropriate to represent better the situation of the system when data are not known with certainty, or a level of uncertainty affects some data. A multi-scenario multi-commodity time-extended optimization model is shown to minimize inventory, handling and transportation costs while meeting demand and supply requirements in every port of Asia. The method is a multi-scenario optimization model, in which scenarios are linked by non-anticipativity conditions. It is able to provide robust solutions to face the uncertainty affecting crucial parameters both from quantitative and qualitative points of view. The experimentation shows that this method provides high demand fulfillment percentages with respect to different future evolutions. To consider the different possible realizations of uncertain parameters, the paper generates a set of scenarios. These scenarios are collected in an overall mathematical model linked by non-anticipativity constraints to guarantee that decision variables are identical up to time θ and current decisions do can not on the values taken by uncertain parameters. A weight is assigned to each scenario to characterize its relative importance.

The following notation is adopted to introduce the multi-scenario formulation in a compact form:

Sets

- G: Set of scenarios. A single scenario g represents a possible future evolution based on the value taken by the uncertain parameters.

Data

- θ: Period up to which all parameters are certain. θ + 1 represents the first period in which uncertain parameters may appear.

w_g Weight of scenario $g \in G$;

The multi-scenario can be represented in the following compact form:

$$\sum_{k=0}^{n} \binom{n}{k} w^g c x^g \qquad (1)$$

$$Ax^g = b_g \ \forall g \in G \qquad (2)$$

$$x_t^g = x_t^f \qquad \forall t \in \{1, \ldots, \theta\}, \forall g, f \in G, g \neq f \quad (3)$$

$$x^g \in Z^+ \cup 0 \quad \forall g \in G \qquad (4)$$

The objective function (1) is weighted over all scenarios. Constraint (1) replicates constraint (2) for each scenario. Constraint (3) can represents the non-anticipativity conditions, enforcing decisions to be identical up to time θ for each pair of scenarios. Equation (4) shows that all variables are non-negative and integer.

It is important to note that the non-anticipativity constrains do not concern both shortage and extra capacity variables. In fact, these information may differ between scenarios since the value taken by the uncertain parameters is different.

The paper presents the multi-scenario formulation in a compact form. The compact form is provided for the deterministic single-scenario formulation because of the sake of clarity. As scenario g represents a deterministic problem instance, the integer programming model could be expressed as below.

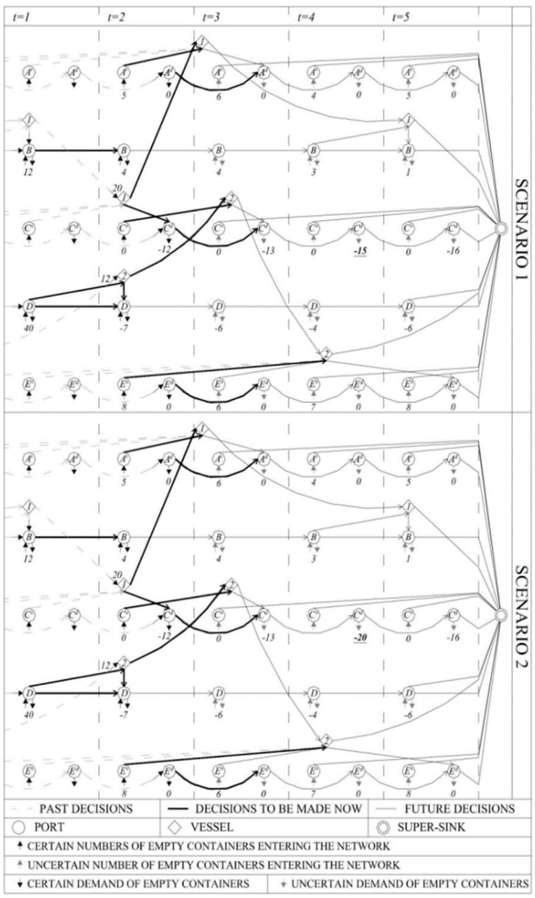


Figure 4: Two scenarios with known and uncertain parameters

Figure 4 shows an example in which uncertainty concerns the numbers of empty containers offered or demanded in ports of Asia. The uncertainty such as Covid 19 pandemic can affect all parameters of the system. The above figure is made up of five ports and two vessels. In each period, the paper replicates the ports, represented by circular-shaped nodes, denoted by letters from A to E. To take into account the restrictions imposed by ports A, C, and E, they are modeled by a pair of nodes in each period. Therefore, the research denotes port A by As and Ad , port C by Cs and Cd , and port E by Es and Ed. Nodes As , Cs and Es are associated with empty containers arriving from the landside, whereas nodes Ad , Cd , and Ed are related to empty containers unloaded from vessels and past inventories. Nodes B and D are associated with containers arriving from both the seaside and the landside. The two vessels are shown by rhombus shaped nodes above the ports where they berth, denoted by numbers 1 and 2. Vessel 1 arrives at port B in period 1, at port C in period 2, at port A in period 3 and then at B in period 5, from where the transportation route on this line is repeated. It means that this vessel will berth in port C in period 6, in port A in period 7, and so on. Vessel 2 arrives at port D in period 2, at port C in period 3, at port E in period 4, it stays at period 5 in this port, from where it repeats the route. Negative numbers represent their demand, whereas positive numbers close to nodes represent the surplus of empty containers. The two scenario in the above figure differs in the number of empty containers requested in port C at period 4, respectively 15 for scenario 1 and 20 for scenario 2.

The benefits of multi-scenario policies could be shown in the context of a regional-size network. For larger cases, ports should be grouped into regions. Furthermore, repositioning problems could be geographically decomposed in order to make this approach tractable.

Innovations facilitate and technology development the development of organizational solutions, intra-channel solutions, and inter-channel solutions. Besides, technological innovations could offer a mathematical model to the ECR problems, contributing directly to the cost reduction of the empty container transportation. One of the most important solutions at present is to free up containers of imported goods and to move the containers in the most reasonable and fast manner to have more containers for shipping export goods. For example, the technology process will optimize the use of containers by moving empty containers from importers directly to exporters and to flexible locations according to regional needs. That will create favorable conditions for businesses to have empty containers for shipping export goods. Rather than bringing back an empty container to the port or assigned depot, transport companies can request to reuse the container and bring it straight to their export customer to get loaded. Other aspects of technological innovations, for example, using more efficient quay cranes and new materials to constructing containers could also contribute to the cost reduction of ECR.

On the other hand, the logistics companies can use methodology to transfer the container lease and the related documentation without bringing the container back to the depot or the terminal and display status information about containers, such as their characteristics, location, and availability.

4. Conclusion

This paper addresses the empty container repositioning problem at a macro scale from a logistics company or an ocean carriers' viewpoint. It also proposes a mathematical model based on multi-scenario learning for the ECR problem as well as suggests solution methodology to solve empty container shortage in Asia. The model can effectively utilize the local information of the shipping network without relying on the accurate prediction for the future, and achieve efficient scheduling by balancing the competition and cooperation between the various agents. The study is still in progress to extend the applicability of this framework. Important attention will be devoted to the development of specialized solution methods to solve large-scale multi-scenario problem instances. One future research endeavor could be to integrate the loaded and empty containers flow decisions in a single integrated model.

Reference

Bandeira DL, Becker JL, Borenstein D. (2009) *A DSS for integrated distribution of empty and full containers*, Decis Support Syst., Elsevier BV, 47(4).

Drewry (2021), World Container Index - 11 Mar https://www.drewry.co.uk/supplychain-advisors/supply-chain-expertise/world-container-index-assessed-by-drewry.

Du Y, Hall R. (1997) *Fleet Sizing and Empty Equipment Redistribution for Center-Terminal Transportation Networks*, Manag Sci. Institute for Operations Research and the Management Sciences (INFORMS), 43(2):145–57. doi.org/10.1287/mnsc.43.2.145.

Firm, S. (2020). *Impact Of COVID-19 On The Shipping And Maritime Industry – Transport – United Arab Emirates*. Retrieved June 18, 2021, from https://www.mondaq.com/marine-shipping/958770/impact-of-covid-19-on-the-shipping-and-maritime-industry.

Li J-A, Liu K, Leung SCH, Lai KK. (2004*) Empty container management in a port with long-run average criterion. Math Comput Model*, Elsevier BV, 40(1-2):85–100. doi.org/10.1016/j.mcm.2003.12.005.

Mehmet Yasin Göçen (2020) *Optimization of costs in empty container repositioning*, ICPR1 2019, LNME, pp. 732–747, 2020.

Myung Y-S, Moon I. (2014) *A network flow model for the optimal allocation of both foldable and standard containers*, Oper Res Lett. Elsevier BV, 42(6-7):484–8. doi.org/10.1016/j.orl.2014.08.004.

Song D-P, Dong J-X. (2012) *Cargo routing and empty container repositioning in multiple shipping service routes.* Transp Res Part B: Method. Elsevier BV, 46(10):1556–75. doi.org/10.1016/j.trb.2012.08.003.

Theofanis, S. and Boile, M. (2009) *Empty marine container logistics: facts, issues and management strategies*, GeoJournal, 74(1), pp.51–65.

UNCTAD (2020), *COVID-19 and maritime transport: Impact and responses*.

UNCTAD (2021), *Review of maritime transport 2020*.

Wang K, Wang S, Zhen L, Qu X. (2014) *Ship type decision considering empty container repositioning and foldable containers*, Transp Res Part E: Logist and Transp Rev. Elsevier BV, 108:97–121. doi.org/10.1016/j.tre.2017.10.003.

DYNAMIC INTEGRATION OF PROCESS PLANNING AND SCHEDULING WITH DUE DATE ASSIGNMENT

Le Thuc Dan Trinh[1], *Nguyen Hoang Huy[2], Duong Vo Nhi Anh[3]

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: [1]ltdtrinh.iuise@gmail.com, [2]nhhuy@hcmiu.edu.vn, [3]dvnanh@hcmiu.edu.vn

Abstract: As considering among alternative process plans gives a high potential to enhance the production efficiency of manufacturing systems, researchers have investigated the field of integration between decision-making stages including scheduling with due date assignment (SWDDA) and the integrated process planning and scheduling (IPPS). Considering as a potential development from the advantage of SWDDA and IPPS, so far, there are just a few studies working on IPPS with Due Date Assignment (IPPSDDA). In this study, the dynamic integrated process planning, scheduling, and due date assignment (DIPPSDDA) is implemented to solve a job shop scheduling problem with random job's arrival time and fixed period maintenance activity. The objective function is to minimize the raising cost caused by earliness, tardiness, and determine due dates for each job. The metaheuristic algorithm, genetic algorithm (GA), has been developed and improved by its hybrid with Tabu Algorithm (TA). Four shop floors different in size have been generated. The performance comparisons of the genetic algorithm (GA) at each shop floor show the efficiency and effectiveness of combining TA into the solving process. In conclusion, computational results show that the proposed combination algorithm (GA/TA) is competitive, give better results than pure GA, and can effectively generate good solutions for DIPPSDDA problems.

Keywords: Dynamic integrated process planning, scheduling, and due date assignment (DIPPSDDA), Genetic Algorithm, Tabu Algorithm.

1. Introduction

1.1. Background Overview

Job shop production often faces many uncertainties which would seriously impact the shop performance if the system is not robust enough. Especially, in the industry of highly customized production where each product is unique and needed to go through a differently determined routing and sequences of operations, it is crucial to smooth the production decision-making process. The commonly desirable objective is to flexibly respond to dynamic events such as machine breakdowns or randomly job arriving and at the same time cope with process-inherent uncertainty and limited resources. Strengthen the flexibility and effectiveness of the production system will help manufacturers thrive in this competitive industry. Besides investing more in modern technology and machines to exceed the factory's capacity, many researches have been focused on innovating new decisions making system to minimize uncertainty impacts and maximize flexibility from the planning to scheduling stage.

1.2. Problem statement

By In this paper, we narrow down the decision-making scheme to three main processes which directly affect the production performance:

- Assuming all jobs are accepted, the route of operations and machines are feasibly and economically decided among alternative routes at the process planning stage. The predetermined process plan (sequence of operations for each job) is usually generated regardless of other jobs' sequence. This may lead to fixed process plans would burden scheduling planner, favoring resources are repeatedly chosen, and they may cause serious bottlenecks or even infeasibility when it comes to practice in the next stage. Besides that, fixed process plans would burden scheduling planner to balance resource load and optimize total shop floor performance.
- Basing on the recent manufacturing condition and required sequenced processing of machines, a due date is assigned. It is challenging to make an independent decision for due dates in a job shop system where uncertainty occurs frequently. The due dates made in the earlier stage later can become less efficient or even lead to high penalty relating to earliness, tardiness in orders' delivery.
- The job is released to the shop floor and machines schedule is provided. This would be the most burdened and challenged stage, for the traditional decision-making strategy, as the fixed two earlier stages fence up choices. Setting due date for jobs

beforehand might cause unnecessary buffer allowance due to uncertainty in production.

Recently, researchers have proved that the highest level of integration (IPPSDDA) gives the best performance on minimizing earliness, tardiness, and due date penalties, and outperforms the lower integration level. However, it is challenging to get the sufficiently efficient result in a reasonable computing time. Thus, the problem rising is to study this approach of integration all three mentioned stages of decision-making scheme so that the result would be better than others well-known methods, and at the same time, the required computing time to get that result is reasonable and beneficial.

1.3. Objectives

The foremost objective of this paper is to successfully adopt the DIPPSDDA approach to cause relaxation and simplification in some aspect of a study process. Besides, this paper aims to provide an additional approach to DIPPSDDA so that it is more applicable in the real-life system than the previous approaches on the field (in terms of adding practical aspect in production that is a preventive maintenance activity scheme).

2. Literature review

2.1. Integrated Process Planning and Scheduling (IPPS) problems

Conventionally, process planning and scheduling functions are treated separately and sequentially where only optimal process plans transferred to the scheduling stage. With the objective to make the system more flexible during production, the idea of integrating process planning with scheduling has arisen such as the study of Yue, G., Tailai, G., & Dan, W. (2021). The impacts of alternative process plans on manufacturing performance is proposed by (Usher, 2003), and the availability of application and their effects on manufacturing system performance of alternative process plans are studied by (Li et al., 2012). CAPP facilitates the process planning process and makes IPPS problem easier than before. There are several literatures studied the integration of CAPP and scheduling such as: (Usher & Femandes, 2012), (Aldakhilallah & Ramesh, 1999), and (Kumar & Rajotia, 2003).

Since applying mathematical techniques is not practical for large-scale problems artificial intelligent techniques such as neural network, meta-heuristic algorithm, multi-agents and evolutionary algorithms have been utilized to solve the IPPS problem with sufficiently effective solution in reasonable computing times (Demir et al., 2014). (Erden et al., 2019) the study has classified these solution approach as Multi-Agent Systems (Li et al., 2010), (Wong et al., 2006), Evolutionary Algorithm (Li et al., 2010), (Gen et al., 2009), (Moon & Seo, 2005), Game Theory (Li et al., 2012), Genetic Algorithm (H. Xia et al., 2016), (Zhang & Wong, 2015), Honey Bee Optimization, Imperialist Competitive Algorithm, Particle Swarm Optimization, Simulated Annealing, Hybrid Approaches and Ant Colony Optimization.

2.2. Scheduling with Due Date Assignment (SWDDA) problems

Static due-date setting methods such as using constant allowance (CON), number of operations in a job (NOPS), total work content plus a constant slack (SLK) and allowances proportional to total work content (TWK) as flow time predictors are studied in early researches (Baker & Bertrand, 1981), (Baker, 1984). (Ragatz & Mabert, 1984) provided a conceptual model of due-date management in job shops. Recent studies focus in determining due-dates based on dynamically changing shop conditions (i.e. dynamic due-date).

According to the number of machines scheduled, we can observe works in two groups: Single machine scheduling with due date assignment (SMSWDDA) and Multi-machine scheduling with due date assignment (MMSWDDA). With single machine problem, in 1999, (V. S. Gordon & Strusevich, 1999) investigates setting due date by a constant slack (SLK) and scheduling n jobs for a machine to minimize earliness penalties. A schedule is only feasible when there are no tardy jobs and precedence constraints of jobs were also concluded. In 2006, (Wang, 2006) proposes a common due-date assignment (CON) in production of controllable processing time. The objective is to minimize cost associated with the common due date, processing time as well as job absolute value in lateness. Also studying on single machine scheduling, processing time dependence on its position in a processing sequence is studied by (Valery S. Gordon & Strusevich, 2009). (Y. Xia et al., 2008) introduces an optimal function for jobs' due date based on the system characteristics and objective function. Considering Just-In-Time (JIT) environment, a heuristic procedure was developed to find job sequence and due date assignment to minimize a linear combination of three penalties: penalty on job earliness, penalty on job tardiness, and penalty associated with long due date assignment. There are numerous SWDDA studies aiming to minimize Earliness/Tardiness cost (E/T) that can be found in (Qi et al., 2002), (Biskup & Jahnke, 2001) most of them conducting CON or SLK due date assignment method.

2.3. Dynamic job shop scheduling – related studies

In the review literature given by Jatoth Mohan, Krishnanand Lanka and A. (Mohan et al., 2019), dynamic events (i.e. stochastic events) in dynamic scheduling are classified into four main groups:

- Workpiece related events: uncertain processing time, the workpiece arrives randomly, the delivery date changes, and dynamic priority and order are changing.

- Machine-related events: machine deterioration, machine breakdowns.
- Process-related events: process delayed, quality rejected and production unstable.
- Other events: the absence of the operating person, the late arrival of raw materials, raw materials defects, etc.

Predict reactive scheduling is common strategy concluding two steps: 1) Generate pre-scheduling scheme without considering the dynamic events in the future, 2) Update the schedule in respond to dynamic events occurrence. According to researchers, the predict reactive scheduling can generate better system performance as increasing productivity and minimizing operating costs compared with totally reactive scheduling (Mohan et al., 2019).

In totally reactive scheduling, schedules are easily generated using dispatching rules (i.e. scheduling rules). Different dispatching rules have different priority determination regarding to various aspects in production. For example, the simplest rule is First In First Out (FIFO) rule stating that jobs are processed in the same sequence as their arrival to a machine. All the commonly used rules are listed in the Table 1.

Table 13. List of common dispatching rules in dynamic scheduling method

No.	Name	Explanation
1	WATC	Weighted Apparent Tardiness Cost
2	ATC	Apparent Tardiness Cost
3	WMS	Weighted Minimum Slack
4	MS	Minimum Slack
5	WSPT	Weighted Shortest Process Time
6	SPT	Shortest Process Time
7	WLPT	Weighted Longest Process Time
8	LPT	Longest Process Time
9	WSOT	Weighted Shortest Operation Time
10	SOT	Shortest Operation Time
11	WLOT	Weighted Longest Operation Time
12	LOT	Longest Operation Time
13	EDD	Earliest Due Date
14	WEDD	Weighted Earliest Due Date
15	ERD	Earliest Release Date
16	WERD	Weighted Earliest Release Date
17	SIRO	Service In Random Order
18	FIFO	First In First Out
19	LIFO	Last In First Out

2.4. Integrated Process Planning and Scheduling with Due Date Assignment (IPPSDDA) problems

The definition of the Integrated Process Planning and Scheduling with Due Date Assignment (IPPSDDA) was introduced by Halil İbrahim Demir et al. for the first time in 2015 (Demir et al., 2014). In the study, process planning, and due-date assignment are simultaneously treated with Apparent Tardiness Cost (ATC) dispatching rule comparing with Service in Random Order (SIRO) rule. Using genetic algorithm (genetic search) and random search, the objective is to embrace JIT philosophy, punish earliness, tardiness and long due date assignment. There are five types of due date assignment considered. Together with the full integration solution, they also run the lower level of integration from totally separate functions to integrated due-date assignment with process plans selection. The authors have concluded that as they increased integration level the solution became better. The genetic search also appears to outperform random search. (Demir & Erden, 2017) continues on the problem with Weighted Earliest Due Date (WEDD) scheduling rule and Weighted Due Date (WDD) Assignment. There are various methods applied and compared: Simulated Annealing (SA), Evolutionary Strategies (ES), Random Search (RS), Hybrid of Random Search and Simulated Annealing (RS/SA), and Hybrid of Random Search and Evolutionary Strategies (RS/ES). As expected, a fully integrated combination is found as the best integration level, giving the best improvement in penalty function. Directed search heuristics outperform random search. Besides, combining random search with other searches was found to be promising metaheuristics. In the solution graph, it appeared that ES and RS/ES methods give the better solution throughout 5 sets of data in the full integration problem. By changing the due date assignment rule to Weighted Number of Operations Plus Processing Times (WNOPP), with the same system, the outperforming methods appears as GA, RS/GA and RS/ES (Demir and Erden, 2017).

From 2015 to 2018, the studies of IPPSDDA are focused on proving the need of integration to improve the overall performance of the system in a static environment, without considering on dynamic events such as machine breakdowns, new job arrivals or confliction with maintenance time of machines. For the first time, dynamic scheduling has been studied with job arriving stochastically in 2019 (Wong et al., 2006). The problem now becomes Dynamic Integrated Process Planning, Scheduling, and Due Date Assignment (DIPPSDDA). Besides working on various metaheuristic algorithms including GA, SA, TA, GA/SA, GA/TA and utilizing 23 dispatching and 37 due date assignment rules, they build a discreate event simulation model to adapt with job coming to the shop floor at different time. The objective is. consistently as previous studies, to minimize the penalties from earliness, tardiness and due dates. Based on the tested experiments, the authors concluded that the proposed hybrid algorithm (GA/TA) and TA algorithms have generally a good frame for the integration problem. Besides, GA/TA and TA show higher reliability to solve DIPPSDDA. They suggested the future work to make the number of operations of each job varies, or tailor to the dynamic events such as machine breakdowns, cancellation of jobs, and the arrival of new urgent jobs, or to make multi-objective such as minimizing makespan, wait times of each job, balancing machine workload.

3. Methodology

3.1. Conceptual Design Description

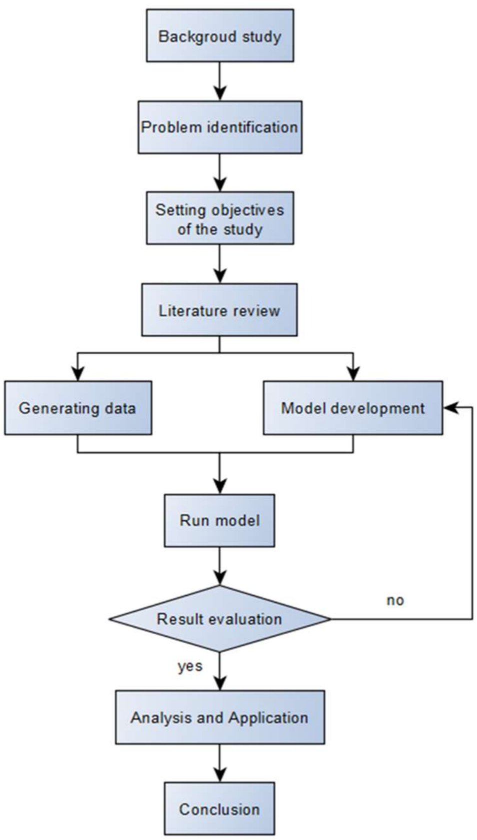


Figure 1. Conceptual design for this study

4. Modeling

4.1 General concept

The idea of the study is to decide three kinds of decision variables simultaneously for the job shop system. They are one dispatching rule, one due date assignment rule and one route for each job to be processed. The study has utilized 36 due date assignment rules and 22 dispatching rules with some predetermined parameters. The combination of decision variables is generated by a search algorithm and it is chosen based on the objective value corresponding to the expected completion time and assigned due date gotten from simulation. The general concept from the study is illustrated in Figure 2.

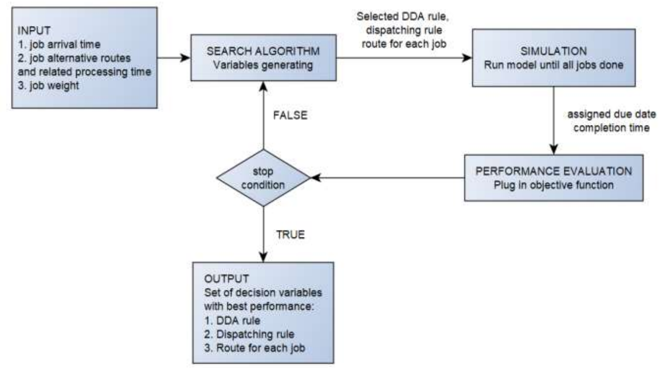


Figure 2. General concept framework by (Erden et al., 2019).

In original approach from the key reference, due date is assigned in the simulation implement, at the time the job arrives at the shop floor, following selected rule. However, the way of accessing the allowance for each rule is solely based on expected waiting time of jobs, regardless to any other unexpected event occurs during production (maintenance activities, machine breakdowns, rework due to quality rejection, etc.).

The approach to calculate due date and related equations (Figure 2) are summarized as below:

Notation in the DDA and dispatching rule:

- d_i : due date time of j-th job
- a_i : the arrival time of j-th job
- p_i : the total processing time of j-th job
- w_i : the weight of the j-th job
- o_j = 10: the number of operations
- P_{av} : the mean processing time of all jobs waiting

The equation for parameters such as the constant q_x and multiplier k_x (x=1,2,3) given as (Wong et al., 2006):

- $q1=P_{av}/2$, $q2=P_{av}$, $q3=3*P_{av}/2$
- k=1, 2, 3

Rule No	Name	Explanation	Equations
0,1,2	SLK	Slack	$d_i = a_i + p_i + q_x$
3,4,5	WSLK	Weighted slack	$d_i = a_i + p_i + w_{1x}q_x$
6,7,8	TWK	Total work content	$d_i = a_i + k_x p_i$
9,10,11	WTWK	Weighted total work content	$d_i = a_i + w_{1x}k_x p_i$
12,13,14	NOPPT	Number of operations plus processing time	$d_i = a_i + p_i + 5k_x o_i$
15,16,17	WNOPPT	The weighted number of operations plus processing time	$d_i = a_i + p_i + 5w_{1x}k_x o_i$
18	RDM	Random-allowance due dates	$d_i = a_i + N(3P_{av}, P_{av})$
19,20,21,22,23,24,25,26,27	PPW	Processing-time-plus-wait	$d_i = a_i + k_x p_i + q_x$
28,29,30,31,32,33,34,35,36	WPPW	Weighted processing-time-plus-wait	$d_i = a_i + w_{2x}k_x p_i + w_{1x}q_x$

Table 2. Due date assignment (DDA) rules and equations

Modifying the study of Erden et al. (2019) mentioned above, to deal with maintenance occurrence during operation, we will decide the due date assignment rule by the end of the simulation, instead of generating it from the search algorithm, by greedy trying all the rules and choosing the best one according to the objective value**.**

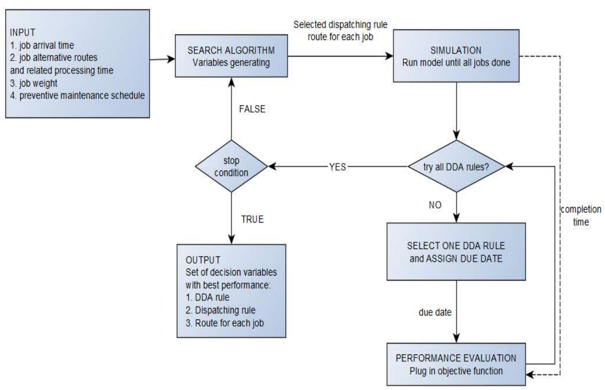


Figure 3. General concept framework (modified)

4.2. Model or Prototype Development

The integrated model in this study can be divided into three main parts. The first part is to implement the search algorithm, construct the set of chosen sequence for each job together with one dispatching rule for the whole process. The second part is to score the performance of generated set by plugging in the experienced completion time from the simulation implemented to the fitness function (objective function). The third part is to learn from the expected completion time of jobs and then choose one rule promising the least cost. The final solution is the feasible plan with the smallest total penalty cost for earliness, tardiness and due date of all jobs.

4.2.1. Initial Model Setting

4.2.1.1. Assumption

Some assumptions of the problem described as following:

- There is no ready job at t = 0, and the arrival time of the job is distributed according to the exponential distribution.
- Due dates for jobs are not assigned at t = 0.
- Each machine can only process a single operation at the same time.
- Queue for waiting job at each machine is unlimited
- Machines are not unexpectedly broken down
- Preventive maintenance routine is carried out every seven and a half hours of operating, and it lasts for thirty minutes. Machine is temporarily stopped during maintenance activity.
- The set-up times of machines are ignored.
- Each job can only be delivered to the customer after all its operations are finished
- There is no rework for jobs and no quality rejected during production
- Arrival time of jobs are determined and unchanged

As attempting to cover the three main decision-making processes which are complicated and extensive in terms of involved information, there might be some aspects that are not mentioned or covered in the solution such as rejecting jobs, due date changing due to negotiation.

4.2.1.2. Objective function

The notation used in the formulation:

- n : number of jobs
- m: number of machines
- d_j : due date time of j-th job
- a_j : the arrival time of j-th job
- p_j : the total processing time of j-th job
- w_j : the weight of the j-th job
- o_j : the number of operations for j-th job
- P_{av} : the mean processing time of all jobs waiting
- p_{ij}: processing time of i-th operation of j-th job
- c_j: completion time of the j-th job
- D_j: the difference between due date and arrival time of j-th job
- E_j: Earliness of j-th job
- T_j: Taridiness of j-th job
- P_D: Penalty for due date
- P_E: Penalty for earliness
- P_T: Penalty for tardiness

Earliness and tardiness are calculated in (1), (2), respectively.

$$E_j = \max(d_j - c_j, 0)$$

$$T_j = \max(c_j - d_j, 0)$$

We consider a working day as two shifts (8 hours/shift) and that makes 960 minutes per day. The penalties for earliness, tardiness and due date are calculated for each day expanse.

$$P_E = w_j \times (5 + 4 \times \frac{E_j}{960})$$

$$P_T = w_j \times (6 + 6 \times \frac{T_j}{960})$$

$$P_D = w_j \times (8 \times \frac{D_j}{960})$$

The total penalty for each job includes penalty for due date, earliness and tardiness:

$$P_{total} = P_D + P_E + P_T$$

The objective is to minimize the overall penalty for all jobs:

$$Minimize\ f = \sum_{j=1}^{n} P_{total\ of\ job\ j}$$

4.2.1.3. Data studied

In this study, three shop floors are learnt and examined with the objective to discover the model's capability and efficiency in improving the decision-making process. The configurations are shown in the Table 3.

	Shop floor 1	Shop floor 2	Shop floor 3
Number of machines	5	15	20
Number of jobs arrived	25	75	100
Number of operations per job (this was fixed at 5 in the key reference)	3 to 5	3 to 15	3 to 20
Number of alternative routes for each job	5	5	5
Number of iterations	150	100	100

Table 3: Parameters for shop floors studied

Processing time (followed normal distribution with the mean time is 6 and the standard deviation is 12 minute) is provided as a technically expected processing time for operations together with scheduled arrival time of jobs to be ready to process at the shop floor.

The data is generated by MATLAB programming. An example of input data for shop floor 1 is shown as arrival_time.xls, route.txt, processing_time.txt in Appendix B, C and D, respectively. The arrival time is illustrated as clock time of arrival (in minute) following exponential distribution with mean time between arrival is 18 minutes. The job processing sequence and corresponding processing time are given following a leading line stated the job index.

4.2.2. Methodology

4.2.2.1. General Methodology

a. Due date assignment rules

There are 37 rules applied for determining due date for jobs as listed in Table 4.2. These rules have been used in many researches recently, and are summarized and adopted by (Wong et al., 2006).

The equation for parameters such as the constant q_x and multiplier k_x (x=1,2,3) given as:

- $q1=P_{av}/2$, $q2=P_{av}$, $q3=3*P_{av}/2$
- k=1, 2, 3

For example, rule number 20 represents the processing-time-plus-wait due date rule, with k=1, $q2=P_{av}$, the equation is:

$$d_j = a_j + p_j + P_{av}$$

Rule No	Name	Explanation	Equations
0,1,2	SLK	Slack	$d_i = a_i + p_i + q_x$
3,4,5	WSLK	Weighted slack	$d_i = a_i + p_i + w_{1x} q_x$
6,7,8	TWK	Total work content	$d_i = a_i + k_x p_i$
9,10,11	WTWK	Weighted total work content	$d_i = a_i + w_{1x} k_x p_i$
12,13,14	NOPPT	Number of operations plus processing time	$d_i = a_i + p_i + 5k_x o_i$
15,16,17	WNOPPT	The weighted number of operations plus processing time	$d_i = a_i + p_i + 5w_{1x} k_x o_i$
18	RDM	Random-allowance due dates	$d_i = a_i + N(3P_{av}, P_{av})$
19,20,21,22,23,24,25,26,27	PPW	Processing-time-plus-wait	$d_i = a_i + k_x p_i + q_x$
28,29,30,31,32,33,34,35,36	WPPW	Weighted processing-time-plus-wait	$d_i = a_i + w_{2x} k_x p_i + w_{1x} q_x$

Table Error! No text of specified style in document.. Due Date Assignment Rules

b. Dispatching rules

In this paper, there are 13 dispatching rules considered to prioritize job to be processed on each machine, each machine will constantly pick the next job, to be processed, that has a greatest priority parameter (I_j) compared with other jobs waiting in queue.

Rule No	Name	Explanation	Equation
1	WSPT	Weighted shortest process time	$I_j = \frac{w_j}{p_j}$
2	SPT	Shortest process time	$I_j = \frac{1}{p_j}$
3	WLPT	Weighted longest process time	$I_j = \frac{p_j}{w_j}$
4	LPT	Longest process time	$I_j = p_j$
5	WSOT	Weighted shortest operation time	$I_j = \frac{w_j}{p_{ij}}$
6	SOT	Shortest operation time	$I_j = \frac{1}{p_{ij}}$
7	WLOT	Weighted longest operation time	$I_j = \frac{p_{ij}}{w_j}$
8	LOT	Longest operation time	$I_j = p_{ij}$
9	ERD	Earliest release date	$I_j = \frac{1}{a_j}$
10	WERD	Weighted earliest release date	$I_j = \frac{w_j}{a_j}$
11	SIRO	Service in random order	$random$
12	FIFO	First in first out	
13	LIFO	Last in first out	

Table 5. Dispatching rules and priority value equations

c. Simulation executed

The simulation model is constructed to simulate the dynamic scheduling characteristic which constantly response to changes (job arrivals).

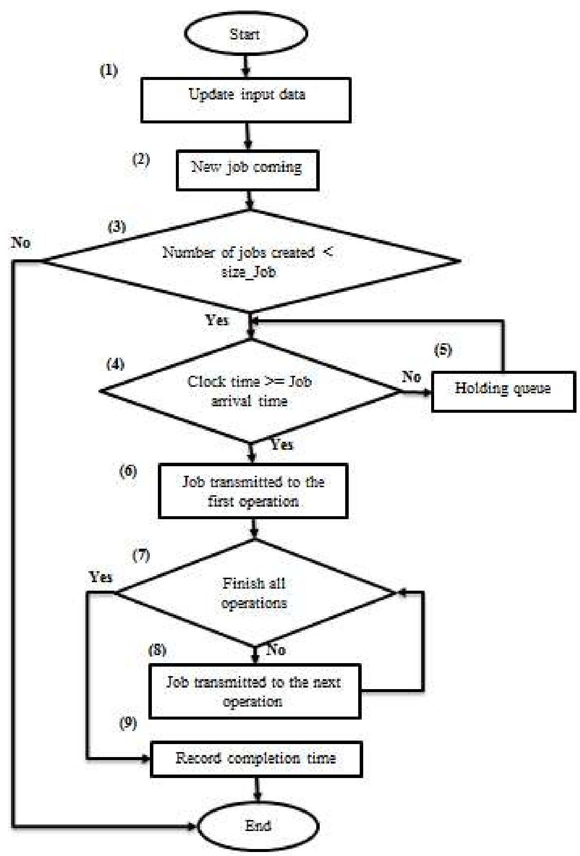


Figure Error! No text of specified style in document.. Manufacturing simulation model's framework

The purpose of running simulation is to get the expected completion time of jobs and then plug into the objective function to examine the performance of such a plan.

In this paper, ARENA simulation application is used together with VBA programming.

4.3. Solution Development

Each shop floor is solved as a separated problem, independent of others. Solving the same data set with GA and GA/TA, the comparison of performance is clearly made in the next chapter.

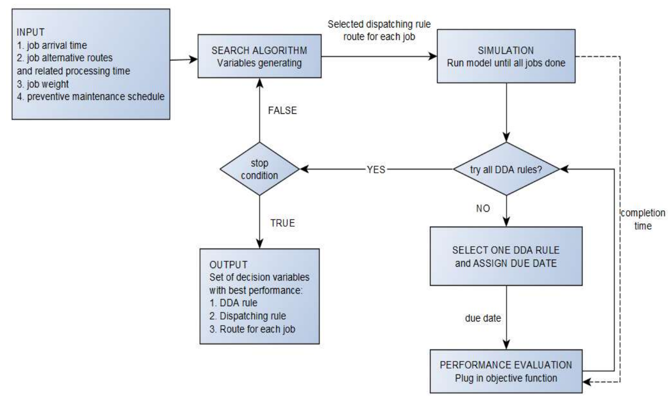


Figure 5. Solution development framework (from input to output)

5. Result Analysis

5.1. Genetic Algorithm Results

As the typical characteristic of metaheuristic technique is its solution good enough but not optimal, the result will vary a bit among different runs.

Since the result's format is the same for every shop floor, below is the result of one time run for shop floor 1 using Genetic Algorithm.

Solution representations (GA – Shop floor 1):

The smallest fitness value is 193.243

The corresponding chromosome is (5, 1, 2, 4, 4, 1, 1, 3, 5, 2, 2, 4, 5, 4, 4, 1, 3, 3, 2, 2, 2, 5, 3, 3, 4, 2)

The first number is the index of dispatching rule 5 (weighted shortest operation time – Table 4.4)

The remaining 25 number is the index of route chosen for 25 jobs

Due date assignment rule is 27 (processing time plus wait – Table 4.3)

5.2. The Hybrid of Genetic Algorithm and Tabu Search Results

Similar to GA, GA/TA also gives slightly different results in each trial. Input the same data as the one presented in the previous section; one time run of GA/TA gives a better result:

Solution representations (GA/TA – Shop floor 1):

The smallest fitness value is 191.634

The corresponding chromosome is (1, 3, 2, 2, 4, 1, 5, 3, 5, 1, 2, 4, 5, 1, 4, 1, 3, 3, 5, 2, 1, 5, 3, 4, 3, 2)

The first number is the index of dispatching rule 5 (weighted shortest process time – Table 4.3)

The remaining 25 number is the index of route chosen for 25 jobs

Due date assignment rule is 31 (weighted processing time plus wait – Table 4.2)

5.3. Validation

Below are the best results of 10 times running GA and 2 times running GA/TA for each shop floor.

	# Ops/ Job	# Jobs	# Iteration	GA		GA/TA			
				Fit	Time (min)	Fit	Time (min)	Saving cost	Time Added (min)
SF 1	3 to 5	25	150	193.243	32	191.634	182	0.83%	150
SF 2	3 to 15	75	100	461.037	25	461.590	128	1.21%	103
SF 3	3 to 20	100	100	631.564	27	623.353	258	1.3%	231

Table 6. Results corresponding to different method of searching

As shown, the GA/TA gives significantly better fitness value for all 3 shop floors from 0.83% to 1.3%. However, the computation time indicates big differences (from more than 1.5 hour to nearly 3.9 hour).

Although 10 times running GA and 2 times running GA/TA for each shop floor are too limited to draw any reliable conclusions on the sensitivity of the methods, there are several important aspects to be considered if the study could be continued and experienced more:

- The slope of improvement in the “best solution so far”, when it seemly reaches local/global optima in both versions (gray line in Figure 5.1 and orange line in Figure 5.2) showing where the solution tending to stop improving or to be cycled in a neighborhood. For meta-heuristics problem, it is significant to spend a majority of time learning the characteristic of the method on the improvement of solution so that adjustment would be done to encourage exploration and better solutions founded throughout iterations. In other words, there is a need to experience deeply and widely into the characteristics of the model, learn its pattern and adjust / navigate searching process (requirement of resources, time dedicated, knowledge on the fields of AI and machine learning, etc.)
- The tradeoffs between better solution and computation time would make the difference in the engineer’s mindset of improving solution. Depending on the industry’s characteristic that the customers will tend to expect fast response on orders (in a competitive market or short-processing time products) or be willing to wait so that there is least risk in lateness, earliness and closest as possible due date for receiving. Based on the company’s strategy, the manager might focus on getting better solution in a specific, acceptable computing time or extending computing time in expectation of improvement.

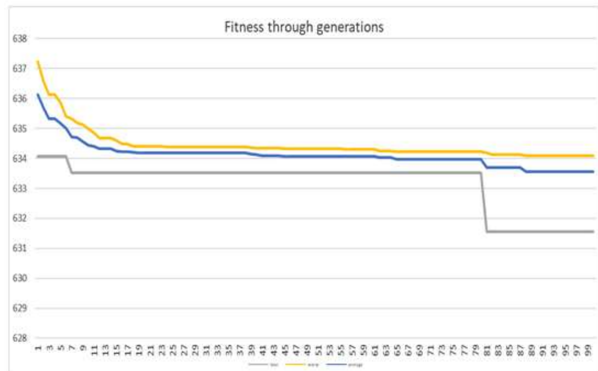


Figure 6. Fitness value through generations using GA (Shop floor 3)

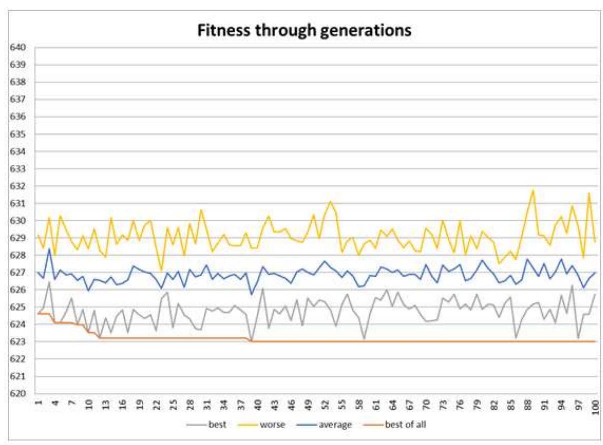


Figure 7. Fitness value through generations using the hybrid GA/TA (Shop floor 3)

Technically, it is notable that running search a large scale on Arena and its old version Visual Basic (1987 – 1999), at some points, makes the results seem to be underrated as its power is not enough to fully activate the proposed method.

Put aside the inability of techs, decision makers should think carefully of which method they use to excellence their work including saving cost, and at the same time satisfying their customers. In reality, deciding on the delivery due date needs to make quick. Despite that the integration attempts definitely show great improvement in efficiency of the whole system in the large view, but to do it fast, organized and flexible is still a big challenge for the technology generation to work towards.

To take benefits of a time consuming solving but giving a promised results, one might need to consider the willing of waiting (for a better deal) of his customers, the potential of other orders canceling or rushed orders jumping in, all the unexpected might happen in the operation like machine breakdowns, etc.

6. Conclusion & Recommendation

6.1. Summary of results

Implementing GA/TA method is challenged in terms of learning deeply the behavior of improving movement so that programmer can increase sensitivity of the method.

Since the number of iterations for Genetic Algorithm is the same but for each iteration, local search is allowed to run at most 5 more times of generating new chromosome and running simulation to get the fitness, the computational time is extremely more expensive. Besides, running the program by old version of Visual Basic (1987 – 1999) embedded in Arena had frequently caused program breakdown and memory interrupted error. In the lack of developing, improving and testing the program, the method might not be really sensitive and ready to handle real-time situations.

Putting aside limitations in implementing and reliable concluding, this study successfully implemented the approach of DIPPSDDA with fixed preventive maintenance activities occurring throughout production using GA and Hybrid GA/TA. Even though there are limited of program running, the best solution (objective value) of GA/TA method showed to outperform the best one of the original GA method (from 0.83% to 1.3% improved). However, computation time is the significant tradeoff needed to be considered in further research (from about 1.5 hours to nearly 3.9 hours more).

6.2. Recommendations for future research

Firstly, running the program by old version of Visual Basic (1987 – 1999) embedded in Arena had frequently caused program breakdown and memory interrupted error. It is highly recommended that other recent modern programming languages, which are capable of running simulation model fast, accurately and allowing developer to access more memory space, should be applied instead.

Secondly, experiencing the behavior or trending in the searching improvement process closely and deeply is also required to explore better solution in a shorter computing time, especially for this type of meta-heuristic which has many strategies and insights of methodology to be considered over searching processes.

Moreover, mixing with lower integration level might be a worthy considered idea to maintain the flexibility in the decision scheme. By regarding the expected performance and the available computing time, managers can decide which level of integration to implement without corrupting recent production progressing in the shop floor.

Reference

Aldakhilallah, K. A., & Ramesh, R. (1999). Computer-Integrated Process Planning and Scheduling (CIPPS): Intelligent support for product design, process planning and control. *International Journal of Production Research*, *37*(3), 481–500. https://doi.org/10.1080/002075499191634

Baker, K. R. (1984). Sequencing Rules and Due-Date Assignments in a Job Shop. *Management Science*, *30*(9), 1093–1104. https://doi.org/10.1287/mnsc.30.9.1093

Baker, K. R., & Bertrand, J. W. M. (1981). An investigation of due-date assignment rules with constrained tightness. *Journal of Operations Management*, *1*(3), 109–120. https://doi.org/10.1016/0272-6963(81)90014-0

Biskup, D., & Jahnke, H. (2001). Common due date assignment for scheduling on a single machine with jointly reducible processing times. *International Journal of Production Economics*, *69*(3), 317–322. https://doi.org/10.1016/S0925-5273(00)00040-2

Demir, H. I., Cakar, T., Ipek, M., Uygun, O., & Sari, M. (2014). Process Planning and Due-date Assignment with ATC Dispatching where Earliness, Tardiness and Due-dates are Punished. *Journal of Industrial and Intelligent Information*, *3*(3), 197–204. https://doi.org/10.12720/jiii.3.3.197-204

Demir, H. I., & Erden, C. (2017). Solving process planning and weighted scheduling with WNOPPT weighted due-date assignment problem using some pure and hybrid meta-heuristics. *Sakarya University Journal of Science*, *21*(2), 1–1. https://doi.org/10.16984/saufenbilder.297014

Erden, C., Demir, H. I., & Kökçam, A. H. (2019). Solving Integrated Process Planning, Dynamic Scheduling, and Due Date Assignment Using Metaheuristic Algorithms. *Mathematical Problems in Engineering*, *2019*. https://doi.org/10.1155/2019/1572614

Gen, M., Lin, L., & Zhang, H. (2009). Evolutionary techniques for optimization problems in integrated manufacturing system: State-of-the-art-survey. *Computers and Industrial Engineering*, *56*(3), 779–

808. https://doi.org/10.1016/j.cie.2008.09.034

Gordon, V. S., & Strusevich, V. A. (1999). Earliness penalties on a single machine subject to precedence constraints: SLK due date assignment. *Computers and Operations Research*, *26*(2), 157–177. https://doi.org/10.1016/S0305-0548(98)00055-0

Gordon, Valery S., & Strusevich, V. A. (2009). Single machine scheduling and due date assignment with positionally dependent processing times. *European Journal of Operational Research*, *198*(1), 57–62. https://doi.org/10.1016/j.ejor.2008.07.044

Kumar, M., & Rajotia, S. (2003). Integration of scheduling with computer aided process planning. *Journal of Materials Processing Technology*, *138*(1–3), 297–300. https://doi.org/10.1016/S0924-0136(03)00088-8

Li, X., Gao, L., & Li, W. (2012). Application of game theory based hybrid algorithm for multi-objective integrated process planning and scheduling. *Expert Systems with Applications*, *39*(1), 288–297. https://doi.org/10.1016/j.eswa.2011.07.019

Li, X., Gao, L., Shao, X., Zhang, C., & Wang, C. (2010). Mathematical modeling and evolutionary algorithm-based approach for integrated process planning and scheduling. *Computers and Operations Research*, *37*(4), 656–667. https://doi.org/10.1016/j.cor.2009.06.008

Mohan, J., Lanka, K., & Rao, A. N. (2019). A review of dynamic job shop scheduling techniques. *Procedia Manufacturing*, *30*, 34–39. https://doi.org/10.1016/j.promfg.2019.02.006

Moon, C., & Seo, Y. (2005). Evolutionary algorithm for advanced process planning and scheduling in a multi-plant. *Computers and Industrial Engineering*, *48*(2), 311–325. https://doi.org/10.1016/j.cie.2005.01.016

Qi, X., Yu, G., & Bard, J. F. (2002). Single machine scheduling with assignable due dates. *Discrete Applied Mathematics*, *122*(1–3), 211–233. https://doi.org/10.1016/S0166-218X(01)00316-X

Ragatz, G. L., & Mabert, V. A. (1984). A framework for the study of due date management in job shops. *International Journal of Production Research*, *22*(4), 685–695. https://doi.org/10.1080/00207548408942488

Usher, J. M. (2003). Evaluating the impact of alternative plans on manufacturing performance. *Computers and Industrial Engineering*, *45*(4), 585–596. https://doi.org/10.1016/S0360-8352(03)00076-7

Usher, J. M., & Femandes, K. J. (2012). Materials Processing Technology. *Advanced Materials Research*, *418–420*(96).

Wang, J. B. (2006). Single machine scheduling with common due date and controllable processing times. *Applied Mathematics and Computation*, *174*(2), 1245–1254. https://doi.org/10.1016/j.amc.2005.05.046

Wong, T. N., Leung, C. W., Mak, K. L., & Fung, R. Y. K. (2006). Integrated process planning and scheduling/rescheduling - An agent-based approach. *International Journal of Production Research*, *44*(18–19), 3627–3655. https://doi.org/10.1080/00207540600675801

Xia, H., Li, X., & Gao, L. (2016). A hybrid genetic algorithm with variable neighborhood search for dynamic integrated process planning and scheduling. *Computers and Industrial Engineering*, *102*, 99–112. https://doi.org/10.1016/j.cie.2016.10.015

Xia, Y., Chen, B., & Yue, J. (2008). Job sequencing and due date assignment in a single machine shop with uncertain processing times. *European Journal of Operational Research*, *184*(1), 63–75. https://doi.org/10.1016/j.ejor.2006.10.058

Zhang, L., & Wong, T. N. (2015). An object-coding genetic algorithm for integrated process planning and scheduling. *European Journal of Operational Research*, *244*(2), 434–444. https://doi.org/10.1016/j.ejor.2015.01.032

Yue, G., Tailai, G., & Dan, W. (2021). Multi-layered coding-based study on optimization algorithms for automobile production logistics scheduling. *Technological Forecasting and Social Change*, *170*, 120889.

DEVELOPMENT OF SERVICES PROVIDERS FOR HO CHI MINH CITY BECOME A LOGISTICS CENTER IN VIETNAM - OBJECTIVES AND SOLUTIONS

Tran Chanh Nghia[1,2*], Mai Thi Dieu[1], Nguyen Chi Khang[1], Nguyen Minh Luat[1] , Dinh Thi Hoai Nhi[1,3] , Do Doan Trang[4], Vo Trong Cang[5]

[1] *Ho Chi Minh city University of Technology (Hutech), Ho Chi Minh city, Viet Nam,*

[2] *cnghiatran1607@gmail.com*

[3] *hoainhi.lgs@gmail.com*

[4] *Binh Duong University, doantrang.bolt@bdu.edu.vn*

[5] *Binh Duong University, vtcang@bdu.edu.vn*

Abstract. Vietnam is considered as a potential and attractive market for the development of the logistics industry. Economic integration has also opened a period of accession to the global common playground with unified regulations and institutions in the field of international trade and services. However, currently, most logistics service providers in Ho Chi Minh City has not been able to grasp this development opportunity. Noticing the above-mentioned problem, many research projects have been proposed to find goals and solutions suitable to the current situation, by analyzing expert opinions, collecting data and studying documents to clarify problems and objectives to be studied. The results were found through media channels such as newspapers and magazines; social networking sites and logistics forums still have a backlog of limitations but are really positive. In addition, the research team also has a few proposals with the hope of contributing to the logistics provider development solution in Ho Chi Minh City.

Keywords: Objectives; Solutions; Service provider; Logistics; Ho Chi Minh City.

1. Overview

Today, people are no stranger to the phrase "Logistics", when it has been proving its role not only for businesses but also for the economy of a country. With its advantages and dynamic, Vietnam has achieved certain successes and attracted the attention of the whole world in the context of the US – China trade war. According to the General Statistics Office of Vietnam, the number of logistics enterprises in Vietnam in the period of 2013 - 2019 has increased sharply, from 1200 enterprises (Thu-Thuy, 2013) in 2013 to 4000 one (Thanh-Chung, 2019) in 2019, proving that the Logistics industry in Vietnam has a great attraction.

The highlight is that up to 70% of logistics enterprises are concentrated in Ho Chi Minh City and it neighborhoods (Thanh-Chung, 2019) due to the advantages of Ho Chi Minh City in terms of source of goods with import and export activities of Ho Chi Minh City and the southern key economic region, and Ho Chi Minh City is located between freight routes from north to south and from east to west, connecting international transport routes through Tan Son Nhat airport as well as port clusters in the south of the country, especially Cat Lai port and Cai Mep – Thi Vai port cluster. It can be said that TP. Ho Chi Minh City is a widely open door for Vietnam to reach out to the world. Therefore, the Logistics development of Ho Chi Minh City plays an important role in the overall development of Vietnam's Logistics industry.

To better explain the objectives and solutions, we have following diagrams as shown in figure 1 and figure 2.

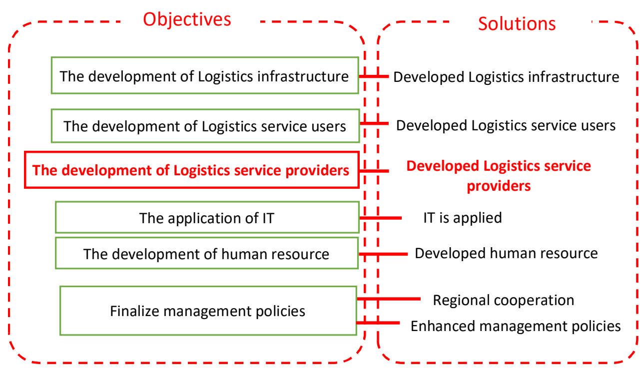


Figure 1. Overall objectives and solutions for the development of Vietnam Logistics industry

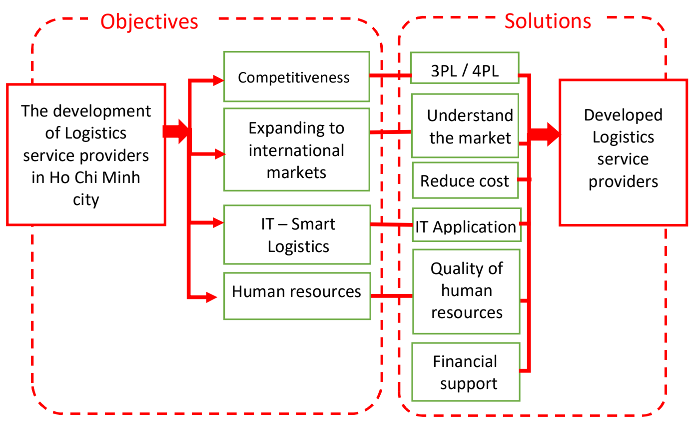


Figure 2. Specific objectives and solutions for the development of Logistics service providers in HCMC

2. Objectives

2.1. Concentrate on improving the competitiveness of Logistics service providers

The goal of developing Logistics service providers in the coming time in Ho Chi Minh City market is promoting the offer of all - inclusive services, connecting supply chains to customers, enhancing the expansion of overseas agents for door-to-door goods delivery to customers. In addition to price competition, the top priority of Logistics companies should be their service quality; meet customers' needs for delivery time, quality of goods guaranteed, error - free paperwork, simple & quick procedures and handle customers' complaints professionally.

2.2. Expanding the customer group of enterprises internationally and deeply participating in supply chain of key industries of Ho Chi Minh City and neighboring provinces

Currently, the growth rate of Vietnam's Logistics industry is approximately about 16 – 20% per year (Vietnam Logistics Business Association – (VLA), 2018). Integrating businesses with diverse services requires close cooperation between enterprises to become professional 3PL and 4PL ones. In addition, companies also set higher targets to raise the outsourcing rate of manufacturing enterprises to over 60%.

2.3. Focusing on the investment of IT applications to improve service delivery efficiency and towards smart logistics

Logistics companies in the coming time should pay attention to investing in advanced I.T applications to improve logistics chain efficiency, such as: Internet of Things, cloud computing, Big Data and block-chain technology that foreign companies currently apply very widely. These technologies promise a lot of benefits, for example, the Internet of Things (IoT) connects all elements of Logistics together in real time, helping businesses respond to incidents almost immediately; and Big Data can analyze a huge amount of stored data, helping us maximize profits, accurately forecast demand, (Vietnam Logistics Research and Development Institute – (VLI), 2019) Therefore, Logistics companies need to make the most of the ability that these technologies bring to offer application products, saving investment costs.

2.4. Improve the quality of human resources at all levels: executive, management and operational; links enterprises with training institutions closely

Professionally train experienced human resources and is fully equipped with professional knowledge, knowledge of domestic and international laws. In order to meet integration age, foreign languages are a strongly important factor; it is necessary to formulate foreign language output standards for students, ensuring that only students who meet foreign language qualifications are considered for graduation.

Additionally, there should be close coordination between the training institutions and businesses, aiming at the goal of students having access to real-life situations while still on their education. Finally, enhancing the quality of the management team, as this is a very important force, contributing to improving the quality of Logistics services.

Above is the analysis of objectives selection from the point of view of the research team. Here we will propose solutions to solve this objectives problem.

3. Solutions

3.1. Developing Logistics companies capable of providing 3PL and 4PL integration services

The potential of small businesses is very tiny, so only joint ventures can develop their capacity. According to the orientation of the Vietnam Association of Logistics Service Enterprises, businesses need to cooperate with each other. Logistics service providers need to rebuild the supply chain for each separate industry, because each industry has its own unique characteristics. For example:

Agro-fishery products need a cold supply chain and are quickly brought to the market to keep their freshness when they reach the customer's table. Currently, the demand for cold supply chains is very large due to the variety of product groups, but the supply is quite limited, only meet 10% of the market demand. Logistics service providers need to develop a cold chain model with standards for certain goods. For wooden household goods, there is no need for a fast shipment, but high precision and care when transporting are required.

That is an example of a number of industries that are growing with strong support from logistics services. Enterprises need to standardize the logistics process from the beginning, clearly stipulate the obligations and responsibilities of each individual involved in the process of providing services.

3.2. Market developing

Associated Logistics Companies: Ho Chi Minh City urgently needs to mobilize seaport enterprises, shipping lines and logistics companies in the area to organize business links with the operators of ICD, warehouses and transportation service providers in order to build a chain of services that support shippers in transporting, storing, and distributing products.

Prioritizing trade, large-scale import and export: Conducting promotional activities calling on commercial traders, large importers and exporters to study solutions to organize freight forwarding at ports and Logistics centers in Ho Chi Minh City.

Proactively attract Authorized Economic Operator (AEO): to transship goods through Ho Chi Minh City ports or to move investment into the desired area, improving the level of customs compliance, security, safety and quality of goods traded with the EU market, taking advantage of large quantities (due to large ships), thus providing fast, quality and affordable services.

3.3. Financial support

The tax incentive program will apply preferential tax rates to attract ship owners, shipping companies, agents and other related companies operating in the territory of Vietnam, as well as companies providing freight and logistics services in Ho Chi Minh City. International shipping companies; shipping agents; forwarding service companies with a global network using Ho Chi Minh City's infrastructure and committed to expanding their operations in Ho Chi Minh City will greatly benefited.

Benefits: Prioritized companies will be tax exempted for 10 years or enjoy a preferential tax rate of no more than 10% for 5 years. The extension will be subject to certain conditions.

Also, the preferential loan program for carriers and container owners will encourage companies to use preferential lending services to rent and buy vehicles and containers. Specifically, companies leasing ships or containers and investment funds enjoy preferential tax of up to 5 years on service revenue; enjoy a tax rate of less than 10% for corporate income tax. Companies with good business results, clear plans and a commitment to provide/use financial support services for ships and containers.

3.4. Enhancing the quality of human resources

Logistics vocational training policies, skills courses such as customs declaration, warehouse operation, ... should be strongly invested in, because the situation of "excess teachers, shortage of workers" is a very serious problem of Vietnamese human resources in general. The lecturers need to be fully equipped with in-depth knowledge and experience in the field of Logistics to have enough equipment to guide students.

It is necessary to invest in building online Logistics training platforms (e-platforms) so that students can research and self-study. Basic and advanced logistics training courses should be organized more often so that the enterprises managers can improve their workmanship and management skill.

3.5. Focusing on the investment of IT application

The Ho Chi Minh City Government should set up a digital map of Logistics points to serve the need of search, statistics, search and Logistics planning in the city area.

Investing in IT application to improve operational efficiency. Towards smart logistics on technology platform 4.0; building technology research and application centers for Logistics and supply chains: developing solutions with low cost and high application range to provide widely to the business community.

3.6. Reducing the cost of Logistics

One of the most costly costs for businesses is tolls, road maintenance and hidden costs such as “tips” for workers or “tips” for customs declaration. Cutting these costs down will help businesses save a fortune to reduce product cost.

Improve the efficiency of management and use of investment capital in infrastructure construction and transport infrastructure development; well exploit the existing transport infrastructure; strengthen the connection between modes of transport, prioritize investment in construction and upgrading of important and urgent works.

Focus on improving logistics infrastructure associated with e-commerce, combining Logistics with e-commerce according to current development trends in the world and the region. Through activities to improve business capacity, we need to focus on implementing Logistics activities on IT platforms and new technologies in Logistics.

The use of software and optimized platforms will help domestic Logistics companies in perform warehouse management, transport management and connection, and delivery management activities more effectively.

Finally, there should be researches and implementation of business cooperation models of Logistics services in Vietnam, including the use of free, easy-to-install applications on mobile devices to support cooperation in order to exploit and utilize the facilities of enterprises. (Nguyen-Hong-Van & Nguyen-Thi-Thuy-Hong, 2021)

4. Judgments and Conclusions

4.1. Judgments on the situation after implementation

Since the implementation of Decision (200/QĐ-TTg, 2017), the logistics industry in general has made significant progress. In early 2021, the government issued Decision (221/QĐ-TTg, 2021) to amend and reaffirm the target by 2025, the proportion of Logistics services' contribution to GDP reached 5% - 6% (compared to 8% - 10% GDP contribution in Decision 200) due to the impact of Covid – 19; the remaining goals are basically achieved. Vietnam's LPI index in 2018 increased by 25 places, ranking 39/160 compared to 64/160 of 2016 (WB, 2018).

However, there are some following limitations. Firstly, Vietnamese Logistics enterprises have not yet exploited their full potential. Secondly, regulations and procedures are still overlapping, causing time consuming. Thirdly, the cost of Logistics services of Vietnamese enterprises is still high (14.5 to 19.2% of GDP, while ASEAN average is 12.7% (VLA, 2018) due to poor Logistics infrastructure connection and severe container shortage in Q4/2020 and Q1/2021.

4.2. Conclusions

Through the process of studying the topic of the discussion and the results obtained, it can be concluded that although there are still many limitations mentioned above, the solutions offered by VLI are reasonable. Besides, researchers also have the following proposals:

- To promote research and development of supporting technologies, improve the rate of use of domestic logistics services of manufacturing companies; Enterprises need to be more proactive in approaching investment in new technology applications software packages like

Warehouse Management System (WMS) and Transport Management System (TMS).

- Continue to promote administrative procedure reform in the direction of transparency and shorten the process to reduce costs equal to other countries in the region; According to Decision No. 221/QD-TTG of the Prime Minister, to develop Logistics human resources advisable to promote logistics training at the university level and Connect training organizations and logistics enterprises in Vietnam with other organizations foreign training institutions.

Reference

Thanh Thuy (2013). Tiềm năng phát triển dịch vụ logictics rất lớn. *Báo điện tử chính phủ.* Access on: http://baochinhphu.vn/Kinh-te/Tiem-nang-phat-trien-dich-vu-logistics-rat-lon/185913.vgp.

Thanh Chung (2019). 70% Doanh nghiệp Logistics tập trung ở TP HCM và vùng lân cận. *Cổng thông tin điện tử chính phủ.* Access on: https://tphcm.chinhphu.vn/70-doanh-nghiep-logistics-tap-trung-o-tphcm-va-vung-lan-can.

VLA - Vietnam Logistics Business Association (2018). S*ách trắng VLA*, trang 59 (Eng: The VLA Whitebook, page 59).

VLI - Vietnam Logistics Research and Development Institute (2019). *Đề án phát triển ngành Logistics trên địa bàn thành phố đến năm 2025, định hướng đến năm 2030* (Eng: Project for the development of HCM city's Logistics industry until 2025 and the orientation to 2030). 2019.

Nguyễn Hồng Vân và Nguyễn Thị Thúy Hồng (2021). Mô hình hợp tác hiệu quả trong hoạt động kinh doanh dịch vụ logistics ở Việt Nam. *Tạp chí Giao thông vận tải,* tháng *5/ 2021* (Eng: Effective cooperation model in the logistics service business in Vietnam. Transport Magazine, May 2021). Access on: https://media.ex-cdn.com/EXP/media.tapchigiaothong.vn/files/attach/2021/05/31/nguen-hong-van-1445.pdf

QD 200/QĐ-TTg (2017). *Quyết định của Thủ tướng Chính Phủ về việc phê duyệt Kế hoạch hành động nâng cao năng lực cạnh tranh và phát triển dịch vụ logistics Việt Nam đến năm 2025* (Eng: Decision No.: 200/QD-TTg, approval for the action plan for improvement of competitiveness and development of vietnam's logistics services by 2025).

QD 221/QĐ-TTg (2021). *Quyết định của Thủ tướng Chính Phủ về việc phê duyệt Kế hoạch hành động nâng cao năng lực cạnh tranh và phát triển dịch vụ logistics Việt Nam đến năm 2025* (Eng: Decision No.: 221/QD-TTg, approval for the action plan for improvement of competitiveness and development of vietnam's logistics services by 2025).

WB - World Bank (2018). *LPI Global Rankings.* Access on: https://lpi.worldbank.org/international /global.

SUPPLIER EVALUATION AND SELECTION PROCESS FOR TEXTILE SUPPLY CHAIN IN VIETNAM: AN MCDM APPROACH

Viet Tinh Nguyen [1,*], To Tam Ngo [2], Ngoc Nhat Linh Le [2] and Thi My Tran Nguyen [2]

[1]*Faculty of Commerce, Van Lang University, Vietnam, tinh.nv@vlu.edu.vn*

[2]*Faculty of Commerce, Van Lang University*

Abstract. In the market economy, competition is typically due to the difficulty in selecting the most suitable supplier, one that is capable to help a business to develop a profit to the highest value threshold and capable to meet sustainable development features. In addition, this research discusses a wide range of consequences from choosing an effective supplier, including reducing production cost, improving product quality, delivering the product on time, and responding flexibly to customer requirements.. In this research, the multi-criteria group decision-making (MCGDM) approach was proposed to solve supplier selection problems. The authors collected data from ten potential suppliers, and the ten main criteria within contain 15 sub-criteria to define the most effective supplier, which has viewed factors, including financial efficiency guarantee, quality of materials, ability to deliver on time, and the conditioned response to the environment to improve the efficiency of the industry supply chain. With the great advantage of possessing low-cost human resources and great potential, Vietnam is one of the expected bright spots. Based on that, it is shown that the selection of an effective supplier to promote optimal competitiveness in a volatile market is complicated in that decision makers must have an understanding of the qualitative characteristics. and quantification to evaluate the symmetry of the effects of the criteria to achieve the most accurate results. The overall goal of the project is to build a Multicriteria Decision-making (MCDM) model to select green suppliers for apparel supply chains using Fuzzy Analysis Network Process (FANP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) methods. The proposed model is then applied to a real-world case study in Vietnam.

Keywords: Textile industry; Supply chain management; MCDM; Supplier selection; FANP; TOPSIS.

1. Introduction

In the context that the 4th Industrial Revolution is taking place rapidly on a worldwide scale, the advanced technologies of this revolution will open up great opportunities for Vietnamese businesses to improve their processes. Science and technology, thereby improving production and business capacity and competitiveness in the global value chain (Bình, 2017). To really make the most of that opportunity, the process of technological innovation in Vietnamese enterprises needs to be done quickly and efficiently.

However, in fact, the science and technology foundation applied in Vietnamese enterprises is still at a weak level and innovation remains weak (The World Bank, 2014). GDP growth has slowed in a less active international context. Former sources of growth are dwindling, increasing the risk of falling into the "middle income trap". Vietnam will have to rely more on increasing labour productivity through innovation. To do so, Vietnam needs to significantly improve its domestic innovation capacity. But the process of restructuring to export "high-tech" goods and services, which are more complex and have a higher knowledge content, is still going on quite slowly. Being stuck in activities that create little added value has limited the ability to learn technology and enhance creativity (WorldBank, 2014). Therefore, it is very important to decide how to invest in technological innovation in a competitive environment.

The objective of this study is to build a multi-criteria decision-making model that supports the decision-making process of technological innovation of Vietnamese enterprises in a scientific and effective manner. The proposed model is built on the theoretical foundation of the management decision-making model according to the Made in Vietnam Lean Management thinking and from the results of practical research on technological innovation decision-making activities of enterprises. Vietnam in today's enterprises (Minh, 2019).

To survive in today's globally competitive environment, businesses must make timely and sound decisions about the effective use of their scarce resources. The contribution of this study is to propose a multi-criteria decision-making model (MCDM) to assist textile supply chains in choosing the best supplier. The applicability of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and the FANP fuzzy analysis procedure is explored as an effective MCDM tool in solving production decision problems (Chakraborty, 2014).

2. Literature review

Multicriteria decision making model (MCDM) have many different methods. In the first stage, all criteria affecting this process are defined by using the supply chain operations reference model (SCOR) and experts' opinion. The results of this research will assist researchers and decision makers in identifying, adapting and applying appropriate methods to identify the optimal material suppliers in the textile and garment industry.

In this study, the authors proposed a multi-criteria decision-making model for supplier selection in the textile industry in Vietnam. SCOR model helps to build a set of criteria, as a prerequisite for the next stage, using the FAHP model to determine the weight of these criteria. The fuzzy analytical network (FANP) process-based methodology is discussed to address the various decision criteria related to the selection of competitive priorities. WSPAS is then used to rank by selected criteria. Finally, PROMETHEE II provided a ranking of potential suppliers and identified the optimal alternative. Quantitative research also helps with modeling, data and statistical analysis, and produces tangible results (Liao, 2018) (Wang, 2018). Safari et al. (2012) developed an integrated MCDM model for supplier evaluation and selection process. In this research, the weights of the evaluation criteria were determined using Shannon's Entropy, while PROMETHEE was used to rank the potential suppliers in the final stage. Senvar et al. (2014) proposed a multi criteria supplier selection model based on a fuzzyPROMETHEE model. The proposed methodology can be used assist decision makers within supply chains in solving similar selection problems. Dag˘ deviren (2008) integrated an approach based on both AHP and PROMETHEE to solve an equipment selection problem. Sari et al. (2016) presented a plausible solution for complex selection problems, by comparing traditional and non-traditional methods. Wang et al. (1986) applied a hybrid fuzzy analytical hierarchy process and green data envelopment analysis for the sustainable supplier selection process in edible oil production, including the fuzzy analytic network process (FANP) and The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) model for solid waste to energy plant location selection in Vietnam. Wang et al. (2018) proposes an MCDM model that uses a combination of SCOR indicators with the AHP-TOPSIS approach to support the supplier selection process. Wang et al. (2018) introduced a FANP-DEA-based approach to supplier selection processes in the rice supply chain. Wang et al. (2018) introduced a FAHP-DEA-based approach to sustainable supplier selection problems in the edible oil industry. Malek et al. (2017) used the gray relational analysis (GRA) method to develop a new hybrid GRA model for the green supply network assessment problem.

The study has implemented FANP and TOPSIS models to select the most suitable supplier and the implementation using case studies has shown that the proposed model is feasible. However, while the proposed model provides important criteria for supplier selection processes in the textile industry, decision makers can vary the number of criteria to better suit their needs. specific needs and situations of their organization. The objective of this study is to find an effective supplier selection decision support tool for the textile industry in dim environment by combining FANP and TOPSIS methods. The FANP method was chosen because of its advantages over FAHP in complex decision-making problems where there are dependencies between criteria. Furthermore, FANP and TOPSIS methods are also easy to understand and available in many decision support software, which increases the usability of the proposed model.

3. Methodology

3.1. Fuzzy analytic network process (FANP) method

Due to Fuzzy analytic network process (FANP) relative simplicity in comparison with Fuzzy analytical hierarchy process (FAHP), FANP is frequently used as an alternative to FAHP in calculating priority weights from fuzzy comparison matrices.

Let $X = \{x_1, x_2, x_{3,} \dots . x_n\}$ be an object set and $O = \{o_1, o_2, o_{3,} \dots . o_n\}$ be a set of goals. The FANP process takes each object from set X then conduct an extended analysis of each goal (o_i) of the object. As such, the extent analysis values of each object (x_i), v, can be shown as followed:

$$V_{o_i}^1, V_{o_i}^2, \dots, V_{o_i}^m, \quad i = 1,2, \dots, n \qquad (2)$$

where $V_{o_i}^j (j = 1,2, \dots ., m)$ are the TFNs.

The extended analysis process can be shown as followed:

Step 1: Determine the fuzzy synthetic extent value of the i^{th} object as:

$$S_i = \sum_{j=1}^{m} V_{o_i}^{j} \otimes \left[\sum_{i=1}^{n}\sum_{j=1}^{m} V_{o_i}^{j}\right]^{-1} \quad (3)$$

With the fuzzy addition operation of *m* extent analysis values of the object matrix ($\sum_{j=1}^{m} V_{o_i}^{j}$) determined by:

$$\sum_{j=1}^{m} V_{o_i}^{j} = \left(\sum_{j=1}^{m} r_j, \sum_{j=1}^{m} p_j, \sum_{j=1}^{m} q_j,\right). \quad (4)$$

The fuzzy additional operation of $V_{a_i}^{j}(j = 1,2, \dots, m)$ values ($\left[\sum_{i=1}^{n}\sum_{j=1}^{m} V_{o_i}^{j}\right]^{-1}$) are calculated by:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} V_{o_i}^{j} = \left(\sum_{j=1}^{n} r_j, \sum_{j=1}^{n} p_j, \sum_{j=1}^{n} q_j,\right). \quad (5)$$

Then, the inversion of the vector in (5) is determined by:

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} V_{o_i}^{j}\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} r_i}, \frac{1}{\sum_{i=1}^{n} p_i}, \frac{1}{\sum_{i=1}^{n} q_i}\right) \quad (6)$$

Step 2: The degree of possibility of $V_2 = (r_2, p_2, q_2) \geq V_1 = (r_1, p_1, q_1)$ is calculated as:

$$\mathrm{P}(V_1 \geq V_2) = sup_{y \geq x}\left[min(\mu_{V_1}(x),), (\mu_{V_2}(y))\right] \quad (7)$$

Which also be shown as:

$$\mathrm{P}(V_1 \geq V_2) = hgt(V_1 \cap V_2) = \mu_{V_2}(d) = \begin{cases} 1 & if p_2 \geq p_1 \\ 0 & if r_1 \geq r_2 \\ \frac{r_1 - q_2}{(p_2 - q_2) - (p_1 - q_1)} & otherwise \end{cases} \quad (8)$$

where *d* is the ordinate of the highest intersection point D between μ_{v_1} and μ_{v_2}. In order to be able to compare V_1 and V_2, we need to calculate the possibility of $(V_1 \geq V_2)$ and $(V_2 \geq V_1)$.

Step 3: Calculate the degree of the possibility that a convex fuzzy number is greater than c convex fuzzy number, with $V_i (i = 1,2, \dots, c)$ as:

$$\mathrm{P}(V \geq V_1, V_2, \dots, V_k) = P[(V \geq V_1) \; and \; (V \geq V_2)\,] \quad (9)$$

and,

$(V \geq V_c)$ = min P(V≥ V_i), $i = 1,2, \dots, c$

Assuming that $d'(B_i) = minP(S_i \geq S_c)$, for $c = 1,2, \dots n$ and $c \# i$, determine the weighted vector as:

$$W' = (d'(B_1), d'(B_2), \dots d'(B_n))^H, \quad (10)$$

Where A_i are n elements.

Step 4: Calculate the normalized weighted vector:

$$\mathrm{d}(B_i) = \frac{d'(B_i)}{\sum_{i=1}^{n} d'(B_1)} \quad (11)$$

$$\mathrm{W} = (d(B_1), d(B_2), \dots, d(B_n))^H \quad (12)$$

3.2. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

Hwang et al. (1981) first proposed the TOPSIS method. A classic TOPSIS process includes 5 steps as followed:

Step 1: Establish the normalized decision matrix.

$$g_{\mathrm{ij}} = \frac{\mathrm{x_{ij}}}{\sqrt{\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{x_{ij}^2}}} \quad (13)$$

where x_{ij} is the original score of decision matrix and g_{ij} is the normalized score of the decision matrix.

Step 2: Determine the normalized weight matrix:

$$o_{ij} = h_j g_{ij} \quad (14)$$

here h_j represents the weight of the j^{th} criterion.

Step 3: The Positive ideal solution (PIS) matrix and Negative ideal solution (NIS) matrix are defined as:

$$\mathrm{O}^+ = o_1^+, o_2^+, \dots, o_{\mathrm{n}}^+$$
$$O^- = o_1^-, o_2^-, \dots, o_{\mathrm{n}}^-; \quad (15)$$

Step 4: Calculate the performance gap between values of each option using the PIS and NIS matrices.

The distance to PIS of each option is calculated as:

$$D_{\mathrm{i}}^+ = \sqrt{\sum_{\mathrm{j=1}}^{\mathrm{m}}(o_{\mathrm{i}}^+ - o_{\mathrm{ij}})^2} \;;\ \mathrm{i} = 1, 2, \dots, \mathrm{m} \quad (16)$$

The distance to NIS of each option is calculated as:

$$D_i^- = \sqrt{\sum_{j=1}^{m}(o_{ij} - o_i^-)^2} \; ; i = 1, 2, \ldots, m \qquad (17)$$

with D_i^+ as the distance to the PIS and D_i^- as the distance to the NIS for the i^{th} option.

Step 5: The preference value (V_i) of each option is calculated as:

$$V_i = \frac{D_i^-}{D_i^- + D_i^+} \qquad i = 1, 2, ..., m \qquad (18)$$

Finally, the V_i values are used to evaluate and rank the potential options

4. Case Study

Guarnieri discussed a number criteria that the textile industry look into namely a total of seven criteria after consulting a number of research. Along with three additional criteria due to the nature of the textile industry in Vietnam, Tab. 1 displays the total list of criteria considered.

Table 14. List of Criteria in determining textile supplier selection

No.	Criteria
1	Culture environment
2	Energy consumption costs
3	Compliance with legislation
4	Public disclosure of environmental records
5	Environmental management
6	ISO 14001 certification
7	Pollution production
8	Lead time
9	Product quality
10	Location of potential wells

A total of 10 suppliers will considered in this research in order to observe how the values change as the MCDM are applied.

5. Results

The authors will conduct the normalization of criteria based on the experts opinion and literature review. The FANP process will be applied initially in order to determine the overall weight of each criteria. Tab. 2 first shows the fuzzy geometric mean for each criteria.

Table 2. Fuzzy Geometric mean of each criteria

Criteria	Fuzzy Geometric Mean of Each Row		
C1	0.9143	1.3419	1.8612
C2	0.8029	1.1227	1.5259
C3	1.0398	1.4639	1.9565
C4	0.8574	1.1919	1.5646
C5	0.9853	1.3381	1.7960
C6	0.4704	0.6290	0.9145
C7	0.5460	0.7178	1.0129
C8	0.6495	0.8987	1.2396
C9	0.6802	0.9428	1.2760
C10	0.5465	0.7431	1.0934

The criteria are then applied with FANP methodology in order to determine fuzzy weights as shown in Tab. 3.

Table 3. Fuzzy weights of each criteria

Criteria	Fuzzy Weights		
C1	0.0642	0.1292	0.2484
C2	0.0564	0.1081	0.2037
C3	0.0730	0.1409	0.2611
C4	0.0602	0.1147	0.2088
C5	0.0692	0.1288	0.2397
C6	0.0330	0.0605	0.1221
C7	0.0383	0.0691	0.1352
C8	0.0456	0.0865	0.1655
C9	0.0478	0.0907	0.1703
C10	0.0384	0.0715	0.1459

The criteria is then normalized in order to determine the final weights of each criteria (shown in Tab. 4) that will be used to determine the final alternative best suited.

Table 4. Final normalized weight for 10 criteria

Criteria	Normalization
C1	0.1289
C2	0.1074
C3	0.1386
C4	0.1120
C5	0.1277
C6	0.0629
C7	0.0708
C8	0.0868
C9	0.0901
C10	0.0747

The alternatives are now first normalized as the first step of TOPSIS shown in Tab. 5 and Tab. 6.

Table 5. Normalized matrix for Criteria 1-5 for 9 alternatives

	C6	C7	C8	C9	C10
A1	0.3621	0.3030	0.3580	0.1391	0.1328
A2	0.2716	0.2273	0.3580	0.1855	0.3542
A3	0.3621	0.3030	0.1989	0.3246	0.1771
A4	0.3169	0.3409	0.3182	0.4174	0.3100
A5	0.1811	0.3409	0.3580	0.3710	0.3985
A6	0.3169	0.3030	0.1989	0.2782	0.1771
A7	0.4074	0.3409	0.2784	0.1855	0.3985
A8	0.1811	0.3409	0.3580	0.4174	0.3542
A9	0.3169	0.3409	0.3182	0.3710	0.3985
A10	0.3621	0.3030	0.3580	0.3246	0.3100

Table 6. Normalized matrix for Criteria 6-10 for 9 alternatives

	C1	C2	C3	C4	C5
A1	0.4196	0.1547	0.1347	0.2289	0.3343
A2	0.3730	0.2579	0.3143	0.4121	0.2971
A3	0.3730	0.3610	0.4041	0.3663	0.3343
A4	0.2331	0.1547	0.3592	0.0916	0.2971
A5	0.2798	0.3610	0.4041	0.2289	0.2971
A6	0.3730	0.2579	0.2245	0.1831	0.2971
A7	0.0933	0.3610	0.4041	0.3663	0.3343
A8	0.4196	0.4641	0.4041	0.4121	0.3343
A9	0.1865	0.2063	0.1347	0.2747	0.2971
A10	0.2331	0.4126	0.1796	0.4121	0.3343

The alternatives are then multiplied accordingly based on the weights of the criteria and is shown in Tab. 7 and 8.

Table 7: Normalized weighted matrix for Criteria 1-5 for 9 alternatives

	C1	C2	C3	C4	C5
A1	0.0541	0.0166	0.0187	0.0256	0.0427
A2	0.0481	0.0277	0.0436	0.0461	0.0379
A3	0.0481	0.0388	0.0560	0.0410	0.0427
A4	0.0301	0.0166	0.0498	0.0103	0.0379
A5	0.0361	0.0388	0.0560	0.0256	0.0379
A6	0.0481	0.0277	0.0311	0.0205	0.0379
A7	0.0120	0.0388	0.0560	0.0410	0.0427
A8	0.0541	0.0499	0.0560	0.0461	0.0427
A9	0.0240	0.0222	0.0187	0.0308	0.0379
A10	0.0301	0.0443	0.0249	0.0461	0.0427

Table 8: Normalized weighted matrix for Criteria 6-10 for 9 alternatives

	C6	C7	C8	C9	C10
A1	0.0228	0.0215	0.0311	0.0125	0.0099
A2	0.0171	0.0161	0.0311	0.0167	0.0264
A3	0.0228	0.0215	0.0173	0.0293	0.0132
A4	0.0199	0.0241	0.0276	0.0376	0.0231
A5	0.0114	0.0241	0.0311	0.0334	0.0298
A6	0.0199	0.0215	0.0173	0.0251	0.0132
A7	0.0256	0.0241	0.0242	0.0167	0.0298
A8	0.0114	0.0241	0.0311	0.0376	0.0264
A9	0.0199	0.0241	0.0276	0.0334	0.0298
A10	0.0228	0.0215	0.0311	0.0293	0.0231

The final ranking for each alternative after considering all the criteria is then shown in Tab. 9.

Table 9: Final ranking of each alternative after considering all the criteria

Alternative	Si+	Si-	Ci	Rank
A1	0.0629	0.0488	0.4367	9
A2	0.0359	0.0620	0.6333	4
A3	0.0271	0.0679	0.7146	2
A4	0.0559	0.0484	0.4642	8
A5	0.0333	0.0615	0.6486	3
A6	0.0499	0.0442	0.4698	7
A7	0.0490	0.0598	0.5497	6
A8	**0.0146**	**0.0821**	**0.8488**	**1**
A9	0.0582	0.0409	0.4126	10
A10	0.0413	0.0570	0.5797	5

From Tab. 9, option A8 gives the best alternatives for all of the possible criteria and alternative by having the highest overall criteria index. From there the decision maker can decide suitably the best location of the new renewable energy project.

6. Conclusion

From the results analyzed above, it is observed that the FANP and the TOPSIS analysis provided sufficient analysis that would assist decision maker in deciding a suitable supplier from a wide range of alternatives. Further studies implies that more MCDM models can be applied to help provide an in-depth analysis of the study.

Reference:

Bình, N. V. (2017). Viet Nam voi cuoc cach mang cong nghiep lan thu 4. Ha Noi: Dai hoc Kinh te Quoc dan.

Brans, J., Vincke, P., & Mareschal, B. (1986). How to select and how to rank projects: The Promethee method. *European Journal Of Operational Research*, *24*(2), 228-238.

Chakraborty, S., & Zavadskas, E. K. (2014). Applications of WASPAS Method in Manufacturing Decision Making. Informatica, pp. 1-20.

Dağdeviren, M. (2008). Decision making in equipment selection: an integrated approach with AHP and PROMETHEE. *Journal Of Intelligent Manufacturing*, *19*(4), 397-406. doi: 10.1007/s10845-008-0091-7

Guarnieri, P., & Trojan, F. (2019). Decision making on supplier selection based on social, ethical, and environmental criteria: A study in the textile industry. Resources, Conservation And Recycling, 141, 347-361. https://doi.org/10.1016/j.resconrec.2018.10.023

Hwang, C.L.; Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. New York: Springer-Verlag.

Liao, H., Wu, D., Huang, Y., Ren, P., Xu, Z., & Verma, M. (2018). Green Logistic Provider

Selection with a Hesitant Fuzzy Linguistic Thermodynamic Method Integrating Cumulative Prospect Theory and PROMETHEE. *Sustainability*, *10*(4), 1291.

Malek, A., Ebrahimnejad, S., & Tavakkoli-Moghaddam, R. (2017). An Improved Hybrid Grey Relational Analysis Approach for Green Resilient Supply Chain Network Assessment. *Sustainability*, *9*(8), 1433.

Minh, N. D. (2019). Xây dựng mô hình ra quyết định đổi mới công nghệ cho doanh nghiệp . Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á, pp. 5-25.

Safari, H., Sadat Fagheyi, M., Sadat Ahangari, S., & Reza Fathi, M. (2012). Applying PROMETHEE Method based on Entropy Weight for Supplier Selection. *Business Management And Strategy*, *3*(1).

Sari, T. and Timor, M. (2016). Integrated supplier selection model using ANP, Taguchi loss function and PROMETHEE methods. *J. Appl. Quant*. Methods, 11, 19–34.

Senvar, O., Tuzkaya, G., & Kahraman, C. (2014). Multi Criteria Supplier Selection Using Fuzzy PROMETHEE Method. *Supply Chain Management Under Fuzziness*, 21-34. doi: 10.1007/978-3-642-53939-8_2

The World Bank. (2014). Đánh giá Khoa học, Công nghệ và Đổi mới sáng tạo ở Việt Nam. Retrieved from WORLDBANK.

Wang, C., Huang, Y., Cheng, I., & Nguyen, V. (2018). A Multi-Criteria Decision-Making (MCDM) Approach Using Hybrid SCOR Metrics, AHP, and TOPSIS for Supplier Evaluation and Selection in the Gas and Oil Industry. *Processes*, *6*(12), 252.

Wang, C., Nguyen, V., Duong, D., & Do, H. (2018). A Hybrid Fuzzy Analytic Network Process (FANP) and Data Envelopment Analysis (DEA) Approach for Supplier Evaluation and Selection in the Rice Supply Chain. *Symmetry*, *10*(6), 221.

Wang, C., Nguyen, V., Thai, H., Tran, N., & Tran, T. (2018). Sustainable Supplier Selection Process in Edible Oil Production by a Hybrid Fuzzy Analytical Hierarchy Process and Green Data Envelopment Analysis for the SMEs Food Processing Industry. *Mathematics*, *6*(12), 302.

Wang, C. N., Nguyen, V. T., Duong, D. H., & Thai, H. T. N. (2018). A hybrid fuzzy analysis network process (FANP) and the technique for order of preference by similarity to ideal solution (TOPSIS) approaches for solid waste to energy plant location selection in Vietnam. Applied Sciences, 8(7), 1100.

WorldBank. (2014). 2014. Gioi thieu bao cao tong quan OECD Nganhang Thegioi.

DISTRIBUTION CENTRE LOCATION SELECTION FOR FMCG SUPPLY CHAINS: A MCDM APPROACH

Van Thanh Nguyen [1,*], Thi Kim Hương Nguyen [2], Thai Thanh Bui [2] and Thanh Sang Thai [2]

[1]*Faculty of Commerce, Van Lang University, Vietnam, thanh.nguyenvan@vlu.edu.vn*

[2]*Faculty of Commerce, Van Lang University*

Abstract. In the context of integration and globalization, where trade liberalization encourage open and interdependent markets, uncertainty has become the norm. With customer increasingly demand shorter lead time and higher service level, combined with increasingly downward pressure on operation costs, fast moving consumer goods (FMCG) supply chains are facing greater challenges than ever before. Therefore, FMCG supply chains are progressively becoming a topic of study. Most FMCG supply networks in Vietnam are still young with more are being developed each year and the problem of choosing the optimal location for a dedicated distribution center is becoming an important and necessary question. Therefore, this study aims to develop a comprehensive MCDM model to support the DC selection process of FMCG supply chains in an uncertain decision-making environment. The proposed model is based on a hybrid approach between SF-AHP and WASPAS and is applied to a case study of fast-growing consumer goods in Vietnam Southern Region.

Keywords: FMCG; Supply chain management; MCDM; Location selection; SF-AHP; WASPAS.

1. Introduction

Fast Moving Consumer Goods in English is Fast Moving Consumer Goods abbreviated as FMCG. FMCG is the fast-moving consumer goods industry, which is defined by products that sell quickly at relatively low costs. This means that FMCG companies rely heavily on profits on the basis of large volumes of goods sold. The higher the number of goods sold and the faster, the more profit. For large manufacturers, hundreds or even thousands of products can be off the production line in a minute. The FMCG industry is generally classified into four sectors as follows: Household appliances, consumer goods, food and beverage, and basic healthcare supplies.

According to a report by Nielsen Vietnam. There was a decrease in sales of major items in the FMCG market such as beer, food, dairy, home/personal care products and tobacco. Only beverage and tobacco products achieved a growth rate of +0.6%, the rest showed signs of decreasing growth rate. The reason for this change is believed to be from the consumer side. Consumers are starting to change their perceptions in their shopping behavior, especially during the holidays.

Together, FMCG industry finds it difficult to adapt to diverse and rapidly changing needs, when distribution is becoming overloaded and slow. FMCG logistics is becoming one of the core competitive advantages of the industry. Location of logistics distribution center has been studied for a very long time. Abundant and various models with algorithms have been proposed in the past. Finding the right locations for the distribution center plays an important role in reducing cost pressure and operating an efficient FMCG supply chain. Choosing the wrong distribution center location can cost your business. Around the world always want to build businesses and choose the optimal location for a dedicated distribution center.

The FMCG supply chain in Vietnam is still young and growing rapidly in the Southern region of Vietnam. And the construction of a distribution center is very necessary, but it is difficult to choose the location of the distribution center. The application of MCDM methods in solving distribution center (DC) selection problems of different supply chains has appeared in many studies.

Therefore, this study aims to develop a comprehensive MCDM model to support the DC selection process of agricultural supply chains in an uncertain decision-making environment. The proposed model is based on a hybrid approach between SF-AHP and WASPAS and is applied to a real case study on FMCG

2. Literature review

The Multi-Criteria Decision Making Method (MCDM) is a very useful tool for daily decision making in various fields. Over the years, many MCDM method-based decision support systems have been developed to assist decision makers in solving complex decision-making problems in various fields, such as science and technology. computer science, environmental science, manufacturing engineering, energy and fuels, etc. Supply chain management is an increasingly popular research topic with many MCDM models being developed to assist in the optimization of various supply chains. Common decision-making issues in the field of supply chain management include choosing a sustainable supplier, choosing a sustainable innovation, choosing a facility location… and many more.

There is much literature on using AHP to solve support distribution center (DC) location selection problems. The analytical fuzzy hierarchical process (AHP) was used by Dobrota et al. (2015) to choose the site of a retail company's distribution center in Serbia. To help with the challenge of picking segmentation locations, Kieu et al. (2021) offer a hybrid MCDM model based on the Spherical Fuzzy Analysis Process (SF-AHP) and the Coherent Compromise Solution (CoCoSo). distribution of agricultural products that are perishable. Basir et al. (2018) utilized AHP and an AHP-based "selection of experts" program software calculator to determine the number of orientations and application distribution center locations a firm should pick in the PT. Unilever Indonesia, TBK branch in Malang. Yang et al. (2015) offer a new technique based on MCDM to address the logistics center site selection process for the FMCG supply chain in China. Kuo et al. (2011) proposed the usage of fuzzy AHP/ANP and fuzzy DEMATEL to aid in the challenge of international DC selection. In order to illustrate the model's and viability, the suggested technique is utilized to tackle the problem of international DC selection in Asia Pacific. Wang et al. (2020) created the hybrid ANP-PROMETHEE II to aid in the evaluation and selection of suppliers in the textile sector. Kaya et al. (2020) determined the priority weights of the criterion, sub-criteria, and alternative locations for the WEEE recycling facility using the Pythagorean fuzzy AHP (PF AHP) technique. Boltürk et al. (2016) offer an Analytic Hierarchy Process (AHP) technique for warehouse location selection in humanitarian logistics (HL) for a Turkish humanitarian assistance organization.

Mihajlovi et al. (2019) evaluated and selected the site of a logistics fruit distribution center in the region using a model (MCDM) combining Analytical Hierarchy Process (AHP) and Assessment of Weighted Total Product (WASPAS) Serbia's southern and eastern regions. (Turskis et al. 2015) propose a fuzzy multi-attribute performance measurement (MAPM) framework to select the best shopping center construction site in Vilnius, combining the advantages of a novel Weighted Aggregated Sum-Product Assessment method with Fuzzy values (WASPAS-F) and Analytical Hierarchy Process (AHP). Yücenur et al. (2021) offers a multi-criteria decision model with four primary criteria, twelve criteria, and three options for site selection for ocean current energy generating plants in Turkey, utilizing the SWARA and WASPAS techniques. Stevi et al. (2018) use the Rough BWM (Worst Method) and Rough WSPAS (Weighed Total Product Assessment) models to find possible roundabout building sites. Stevi et al. (2018) use the Rough BWM (Worst Method) and Rough WSPAS (Weighed Total Product Assessment) models to find possible roundabout building sites.

None of these models are tailored to the problem of DC selection in FMCG supply chains. As a result, the goal of this research is to provide a comprehensive MCDM model to aid in the DC selection process of FMCG supply chains in uncertain decision-making situations. The suggested model is based on a combination of SF-AHP and WASPAS and is applied to a real-world case study in the Mekong Delta region of Viet Nam.

3. Methodology

3.1 Spherical Fuzzy Sets Theory

Multiple MCDM models have variations and one of them involves the spherical fuzzy sets theory. Sharaf (2020) applied spherical fuzzy sets in combination with VIKOR method to solve a supplier selection problem. The application of spherical fuzzy sets assists the decision makers with a larger matrix of preference and opinions. Otay and Atik (2020) create an MCDM model to solve a real-world oil station location evaluation problem using spherical fuzzy sets and the WASPAS method. Sensitivity analysis showed that the proposed model is complex and big. Gül (2020) developed spherical fuzzy extension of the DEMATEL method. The proposed model was applied to a building contractor selection problem.

Spherical fuzzy sets theory was developed recently by Gundogdu and Kahraman (2019) based on Pythagorean fuzzy sets and Neutrosophic sets theories . Pythagorean fuzzy sets' membership functions are defined by membership, non-membership, and hesitancy parameters. While Neutrosophic fuzzy sets membership functions are also composed of truthiness, falsity, and indeterminacy parameters. Spherical fuzzy sets theory is based on the idea that by defining a membership function on a spherical surface, decision makers can generalize different types of fuzzy sets.

The membership function of a spherical fuzzy set is determined by three parameters: the degree of membership, the degree of non-membership, and the degree of hesitancy. Each of these parameters can have an opinionated element between 0 and 1 independently and the sum of the squared values of these parameters is at most 1.

A spherical fuzzy set $\tilde{A}_S$ of the universe of U_1 is defined as:

$$\tilde{A}_S = \{x, (\mu_{\tilde{A}_S}(x), v_{\tilde{A}_S}(x), \pi_{\tilde{A}_S}(x)) | x \epsilon U_1\} \quad (1)$$

with:

$$\mu_{\tilde{A}_S}(x) : U_1 \rightarrow [0,1], v_{\tilde{A}_S}(x) : U_1 \rightarrow [0,1], \text{and } \pi_{\tilde{A}_S}(x) : U_1 \rightarrow [0,1]$$

and

$$0 \leq \mu^2_{\tilde{A}_S}(x) + v^2_{\tilde{A}_S}(x) + \pi^2_{\tilde{A}_S}(x) \leq 1 \quad (2)$$

with $\forall x \in U_1$ and $\mu_{\tilde{A}_S}(x)$ as the degree of membership, $v_{\tilde{A}_S}(x)$ as the degree of non-membership, and $\pi_{\tilde{A}_S}(x)$ as the hesitancy of x to $\tilde{A}_S$.

Basic arithmetic operations - such as union, intersection, addition, multiplication, and power - of spherical fuzzy sets are applied and shown in the work of Gundogdu and Kahraman (2019).

3.2 Spherical Fuzzy Analytic Hierarchy Process (SF-AHP) model

SF-AHP method is introduced by Gundogdu and Kahraman (2019) which is an extension of AHP with spherical fuzzy sets. In this paper, SF-AHP is employed to determine the DC selection criteria weights. The SF-AHP method has 7 steps:

Step 1: Build the model hierarchical structure. A hierarchical structure with three levels is constructed. Level 1 is the goal of the model based on a score index. The score index is determined with *n* criteria which is represented in Level 2 of the structure. A set of *m* alternative A ($m \geq 2$), is defined in Level 3 of the structure.

Step 2: Build pairwise comparison matrices of the criteria using spherical fuzzy judgement based on linguistic terms:

Table 1: Linguistic measures of importance:

Linguistic Measurement	$(\boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\pi})$	**Score Index**
Absolutely more importance (AM)	(0.9, 0.1, 0.0)	9
Very high importance (VH)	(0.8, 0.2, 0.1)	7
High importance (HI)	(0.7, 0.3, 0.2)	5
Slightly more importance (SM)	(0.6, 0.4, 0.3)	3
Equally importance (EI)	(0.5, 0.4, 0.4)	1
Slightly lower importance (SL)	(0.4, 0.6, 0.3)	1/3
Low importance (LI)	(0.3, 0.7, 0.2)	1/5
Very low importance (VL)	(0.2, 0.8, 0.1)	1/7
Absolutely low importance (AL)	(0.1, 0.9, 0.0)	1/9

Equation (3) and (4) are applied to calculate the score indices (SI) of each alternative.

$$\text{SI} = \sqrt{\left|100 * \left[\left(\mu_{\tilde{A}_s} - \pi_{\tilde{A}_s}\right)^2 - \left(v_{\tilde{A}_s} - \pi_{\tilde{A}_s}\right)^2\right]\right|} \quad (3)$$

for AM, VH, HI, SM, and EI.

$$\frac{1}{\text{SI}} = \frac{1}{\sqrt{\left|100 * \left[\left(\mu_{\tilde{A}_s} - \pi_{\tilde{A}_s}\right)^2 - \left(v_{\tilde{A}_s} - \pi_{\tilde{A}_s}\right)^2\right]\right|}} \quad (4)$$

for SL, LI, VL, and AL.

Step 3: Check the consistency of each pairwise comparision matrix.

The classical consistency check is applied with the threhold of the Consitency Ratio (CR) value of 10%:

$$CR = \frac{CI}{RI} \quad (5)$$

With CI as Consistency Index calculated as:

$$CI = \frac{\lambda_{max} - n}{n-1} \quad (6)$$

Where λ_{max} is the maximum eigenvalue of the matrix, and n is the number of criteria.

The Random Index (RI) is determined based on the number of criteria.

Step 4: Obtain the fuzzy weights of criteria and alternatives.

Each alternative's weight with respect to each criterion is obtained using the followed equation:

$$SWM_w(\tilde{A}_{S_{i1}}, \ldots, \tilde{A}_{S_{in}}) = w_1\tilde{A}_{S_{i1}} + \cdots + w_n\tilde{A}_{S_{in}} = \langle\left[1 - \prod_{j=1}^{n}\left(1 - \mu_{\tilde{A}_{S_{ij}}}^2\right)^{w_j}\right]^{0.5}, \prod_{j=1}^{n} V_{\tilde{A}_{S_{ij}}}^{w_j}, \left[\prod_{j=1}^{n}\left(1 - \mu_{\tilde{A}_{S_{ij}}}^2\right)^{w_j} - \prod_{j=1}^{n}\left(1 - \mu_{\tilde{A}_{S_{ij}}}^2 - \pi_{\tilde{A}_{S_{ij}}}^2\right)^{w_j}\right]^{0.5}\rangle \quad (7)$$

where $w = 1/n$.

Step 5: Obtain the global weights using hierarchical layer sequencing.

The final ranking of the alternatives is estimated by aggregating the spherical weights at each level of the hierarchical structure. There are two feasible way to perform the computation at this point.

The first way is using the score function in Equation (8) to defuzzify the criteria weights:

$$S(\tilde{w}_j^S) = \sqrt{\left|100 * \left[\left(3\mu_{\tilde{A}_s} - \frac{\pi_{\tilde{A}_s}}{2}\right)^2 - \left(\frac{v_{\tilde{A}_s}}{2} - \pi_{\tilde{A}_s}\right)^2\right]\right|} \quad (8)$$

Then, the criteria weights are normalized using Equation (9) and apply spherical fuzzy multiplication in Equation (10):

$$\overline{w}_j^s = \frac{S(\tilde{w}_j^S)}{\sum_{j=1}^{n} S(\tilde{w}_j^S)} \quad (9)$$

$$\tilde{A}_{S_{ij}} = \overline{w}_j^s * \tilde{A}_{S_i} = \langle\left(1 - \left(1 - \mu_{\tilde{A}_{S_i}}^2\right)^{\overline{w}_j^s}\right)^{1/2}, v_{\tilde{A}_{S_i}}^{\overline{w}_j^s}, \left(\left(1 - \mu_{\tilde{A}_{S_i}}^2\right)^{\overline{w}_j^s} - \left(1 - \mu_{\tilde{A}_{S_i}}^2 - \pi_{\tilde{A}_{S_i}}^2\right)^{\overline{w}_j^s}\right)^{1/2}\rangle \quad (10)$$

The final ranking score ($\tilde{F}$) for each alternative A_i is calculated using Equation (11):

$$\tilde{F} = \sum_{j=1}^{n} \tilde{A}_{S_{ij}} = \tilde{A}_{S_{i1}} + \tilde{A}_{S_{i2}} + \cdots + \tilde{A}_{S_{in}} \quad (11)$$

Another option is to continue the calculation without the defuzzification of the criteria weights. The spherical fuzzy global weights are calculated as:

$$\prod_{j=1}^{n} \tilde{A}_{S_{ij}} = \tilde{A}_{S_{i1}} * \tilde{A}_{S_{i2}} * \ldots * \tilde{A}_{S_{in}} \quad (12)$$

Then, the final ranking score ($\tilde{F}$) of each alternative is calculated using Equation (11).

3.3 Weighted aggregated sum product assessment (WASPAS)

One of the most utilized and efficient multicriteria decision making models for assessing multiple alternatives in numerous criteria is the Weighted Sum Model (WSM). Firstly, there are a options and b decision criteria. We then define z_b as the importance for the criteria and x_{ab} is the performance level for option a evaluated in criterion b. Finally, the overall relative importance of alternative y, denoted as $P_y^{(1)}$, is defined (Triantaphyllou and Mann, 1989).

$$P_y^{(1)} = \sum_{b=1}^{n} \bar{x}_{ab} z_b \quad (13)$$

Where the linear normalization for each initial criteria value are calculated as follows,

$$\bar{x}_{ab} = \frac{x_{ab}}{max_a x_{ab}} \quad (14)$$

if $max_a x_{ab}$ value is preferable or

$$\bar{x}_{ab} = \frac{min_a x_{ab}}{x_{ab}} \quad (15)$$

if $min_{ab} x_{ab}$ value is preferable.

Another method that is commonly used when assessing multiple alternatives using the total relative importance of option y, denoted as $P_y^{(2)}$ is the Weight Product Model (WPM). It is defined as follows:

$$P_y^{(2)} = \prod_{b=1}^{n} (\bar{x}_{ab})^{z_b} \quad (16)$$

In order in incorporate both methods to evaluate further the importance of the alternatives, the weights of total relative importance are then equally divided between the WSM and WPM results for a total score:

$$P_y = 0.5P_y^{(1)} + 0.5P_y^{(2)} \quad (17)$$

From the study above and evaluating further the accuracy and effectiveness in decision making, the coefficients that defined WSM and WPM can then be further changed to adapt suitably depending on the problem. This change in coefficients in called the Weighted Aggregated Sum Product Assessment method and it is used to rank the alternatives in this research.

$$P_y = \lambda \sum_{b=1}^{n} \bar{x}_{ab} z_b + (1 - \lambda) \prod_{j=1}^{n} (\bar{x}_{ab})^{z_b} \quad (18)$$

4. Case study

A distribution centre location selection for FMCG supply chain case study is performed to demonstrate the model's feasibility. The studied supply chain operates within the Southern regions of Vietnam and is currently consider 5 potential locations for a new DC (A1, A2, A3, A4, and A5).

After consulting with relevant authorities and industries experts, a list of criteria and sub-criteria is identified:

Table 1. Evaluation Criteria

Criteria	Sub-criteria	Criteria	Sub-criteria
Cost	Fixed Cost (C1)	Infrastructure	Land road network (C6)
	Variable Cost (C2)		Available service infrastructures (C7)
Service level	Transportation time (C3)	Sustainability factors	Distance to forest area (C8)
	Distance to main markets (C4)		Distance to surface water (C9)
	Distance to suppliers (C5)		Impact to local communities (C10)

Then, the SF-AHP method is applied to calculate the weights of the sub-criteria:

Table 2. Sub-criteria weights

Criteria	Spherical Fuzzy Weights			Crisp Weights
C1	0.3221	0.6704	0.2362	0.0528
C2	0.3622	0.6480	0.2272	0.0607
C3	0.4008	0.6236	0.2273	0.0681
C4	0.4807	0.5377	0.2322	0.0831
C5	0.5348	0.4930	0.2199	0.0937
C6	0.5936	0.4397	0.1956	0.1056
C7	0.6903	0.3339	0.1608	0.1249
C8	0.7357	0.2797	0.1445	0.1339
C9	0.7712	0.2370	0.1366	0.1409
C10	0.7533	0.2504	0.1764	0.1362

Finally, the WASPAS model is applies to calculate the performance score of each alternative and produce the final ranking. The weighted normalized matrix, exponentially weighted matrix and the final ranking are provided in Table 3, Table 4, and Table 5:

Table 3. Weighted normalized matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
A1	0.053	0.047	0.068	0.074	0.083	0.106	0.097	0.134	0.110	0.136
A2	0.047	0.054	0.061	0.083	0.094	0.094	0.111	0.119	0.110	0.121
A3	0.047	0.054	0.068	0.074	0.083	0.094	0.111	0.134	0.110	0.136
A4	0.053	0.061	0.061	0.083	0.094	0.094	0.125	0.119	0.141	0.106
A5	0.041	0.054	0.053	0.074	0.083	0.106	0.111	0.119	0.125	0.136

Table 4. Exponentially weighted matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
A1	1.000	0.985	1.000	0.990	0.989	1.000	0.969	1.000	0.965	1.000
A2	0.994	0.993	0.992	1.000	1.000	0.988	0.985	0.984	0.965	0.984
A3	0.994	0.993	1.000	0.990	0.989	0.988	0.985	1.000	0.965	1.000
A4	1.000	1.000	0.992	1.000	1.000	0.988	1.000	0.984	1.000	0.966
A5	0.987	0.993	0.983	0.990	0.989	1.000	0.985	0.984	0.984	1.000

Table 5. Final ranking

Alternatives	Qi1	Qi2	Qi	Ranking
A1	0.9078	0.9078	0.9078	3
A2	0.8929	0.8929	0.8929	5
A3	0.9108	0.9108	0.9108	2
A4	**0.9356**	**0.9356**	**0.9356**	**1**
A5	0.9023	0.9023	0.9023	4

The results suggest that location number 4 (A4) is the optimal location, followed by A3, A1, A5, and A2.

5. Conclusion

Facility location selection problems are complex decision-making problems that frequently required the application of MCDM methods to solve. This study aims to contribute to the available methods with a new approach using Spherical Fuzzy Sets theory. The proposed model is an easy-to-perform and comprehensive approach to solving DC location selection problems. The model also has demonstrated its feasibility through the case study. The model can be modified by adding or changing criteria and applied to similar facility location selection problems. In the future, researchers can perform comparative studies to compare the performance of the model and other MCDM model such as FAHP-TOPSIS or FANP-DEMATEL.

Reference

Ayyildiz, E., & Gumus, A. T. (2020). A novel spherical fuzzy AHP-integrated spherical WASPAS methodology for petrol station location selection problem: a real case study for İstanbul. Environmental Science and Pollution Research, 27(29), 36109-36120.

Basir, A. M. (2018). *PENENTUAN LOKASI PUSAT DISTRIBUSI PRODUK FAST MOVING CONSUMER GOODS PADA PT. UNILEVER INDONESIA, TBK CABANG MALANG* (Doctoral dissertation, University Of Muhammadiyah Malang).

Boltürk, E., Çevik Onar, S., Öztayşi, B., Kahraman, C., & Goztepe, K. (2016). Multi-attribute warehouse location selection in humanitarian logistics using hesitant fuzzy AHP. *International Journal of the Analytic Hierarchy Process*, *8*(2), 271-298.

Dobrota, M., Macura, D., & Šelmi, M. (2015). Multi criteria decision making for distribution center location selection-Serbia case study. In *2nd Logistics International Conference* (pp. 32-37).

Gül, S. (2020). Spherical fuzzy extension of DEMATEL (SF-DEMATEL). *International Journal Of Intelligent Systems*, *35*(9), 1329-1353.

Kaya, A., Çiçekalan, B., & Çebi, F. (2020). Location selection for WEEE recycling plant by using Pythagorean fuzzy AHP. *Journal of Intelligent & Fuzzy Systems*, *38*(1), 1097-1106.

Kieu, P. T., Nguyen, V. T., Nguyen, V. T., & Ho, T. P. (2021). A Spherical Fuzzy Analytic Hierarchy Process (SF-AHP) and Combined Compromise Solution (CoCoSo) Algorithm in Distribution Center Location Selection: A Case Study in Agricultural Supply Chain. *Axioms*, *10*(2), 53.

Kuo, M. (2011). Optimal location selection for an international distribution center by using a new hybrid method. *Expert Systems With Applications*, *38*(6), 7208-7221.

Kutlu Gündoğdu, F., & Kahraman, C. (2019). Spherical fuzzy sets and spherical fuzzy TOPSIS method. *Journal Of Intelligent & Fuzzy Systems*, *36*(1), 337-352.

Mihajlović, J., Rajković, P., Petrović, G., & Ćirić, D. (2019). The selection of the logistics distribution center location based on MCDM methodology in southern and eastern region in Serbia. Operational Research in Engineering Sciences: Theory and Applications, 2(2), 72-85.

Otay, I., & Atik, S. (2020). Multi-criteria Oil Station Location Evaluation Using Spherical AHP&WASPAS: A Real-Life Case Study. *Advances In Intelligent Systems And Computing*, 591-598.

Sharaf, I. (2020). Spherical Fuzzy VIKOR with SWAM and SWGM Operators for MCDM. *Decision Making With Spherical Fuzzy Sets*, 217-240.

Smarandache, F. (1999). A unifying field in logics neutrosophy: neutrosophic probability, set and logic. American Research Press, Rehoboth, pp 1–141

Stević, Ž., Pamučar, D., Subotić, M., Antuchevičiene, J., & Zavadskas, E. K. (2018). The location selection for roundabout construction using Rough BWM-Rough WASPAS approach based on a new Rough Hamy aggregator. Sustainability, 10(8), 2817.

Triantaphyllou, E., & Mann, S. H. (1989). An examination of the effectiveness of multi-dimensional decision-making methods: a decision-making paradox. *Decision Support Systems*, *5*(3), 303-312.

Turskis, Z., Zavadskas, E. K., Antucheviciene, J., & Kosareva, N. (2015). A hybrid model based on fuzzy AHP and fuzzy WASPAS for construction site selection. International Journal of Computers communications & control, 10(6), 113-128.

Wang, C., Viet, V., Ho, T., Nguyen, V., & Nguyen, V. (2020). Multi-Criteria Decision Model for the Selection of Suppliers in the Textile Industry. *Symmetry*, *12*(6), 979.

Yager, R. (2013) Pythagorean fuzzy subsets. In 2013 Joint IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS). IEEE, pp 57–61.

Yang, Y., Fan, J., & Zhang, L. (2015). Study on Location Selection of FMCG Logistics Centers Based on Fuzzy MCDM Model. *Applied Mechanics And Materials*, *744-746*, 1888-1894.

Yücenur, G. N., & Ipekçi, A. (2021). SWARA/WASPAS methods for a marine current energy plant location selection problem. Renewable Energy, 163, 1287-1298.

SUPPLIER EVALUATION AND SELECTION PROCESS FOR RENEWABLE ENERGY PROJECTS: AN MCDM APPROACH

Syed Tam Husain [1,*], Tin Giang Cam [2], Minh Thu Nguyen [2] and Hong Vy Vo [2]

[1] *Faculty of Commerce, Van Lang University, Vietnam, tam.husain@vlu.edu.vn*

[2] *Faculty of Commerce, Van Lang University*

Abstract. Renewable Energy (RE) is a clean, inexhaustible, and increasingly competitive source of energy. They do not produce greenhouse gases, climate change and pollute emissions. According to the International Renewable Energy Agency (IRENA), doubling the share of renewables in the world energy mix, to 36% by 2030, would lead to an additional 1.1% in global growth by 2030. That year ($1.3 trillion) benefits increased by 3.7% and the number of jobs in the sector amounted to more than 24 million people, compared with 9.2 million today. Projects on the RE sector such as wind, hydropower and solar power have made certain contributions to the stabilization of the national electricity network in Vietnam. In this study, the authors presented two research methods: the method of dosing and nailing. For the dosing research method, apply the Fuzzy Analytic Network Process (FANP) model to solve the complex problems of decision-making related to the different gendered criteria in supplier selection and use the Data Envelopment Analysis (DEA) model to rank potential suppliers. The contribution of this study is to assist in identifying important criteria in the process of selecting the right suppliers for RE project to meet the needs of sustainable socio-economic development of Vietnam.

Keywords: Renewable energy (RE); Supply chain management; MCDM; Supplier selection; FANP; DEA.

1. Introduction

Nowadays, environmental pollution is the biggest problem in the world. Specifically, the phenomenon of global warming has caused the ices caps to melt of poles. Besides, when the amount of carbon dioxide (CO_2) in the air increases, it leads to the formation of the greenhouse effect, and this also contributes to the destruction of the ozone layer. Moreover, In the areas where the ozone layer is depleted, the soil in that area is gradually deserted, losing the current ecological balance. Therefore, this research proposes to use renewable energy (RE) to replace current forms of energy because generating RE is also decreasing and at a sustainable level, while the overall cost trend for fossil fuels to go in the opposite direction Besides, RE is a clean, environmentally friendly, low-pollution energy source that does not cause greenhouse effects. In addition, RE will never run out, very useful and helping to save electricity for households, factories, factories, etc. RE has five main forms: wind energy, hydroelectric power, geothermal energy, solar energy, and bioenergy. Most of them are a great source of energy from nature, can be used for many different needs and locations. The truth is that supplier selection can happen at any stage of the supply chain, so this research has researched and brought to everyone the Multi-Criteria Decision Making (MCDM) the supplier selection process for RE projects. This research used two process approaches to mixed Fuzzy Analysis Network (FANP) and Data Envelopment Analysis (DEA) to evaluate and select suppliers for energy projects regenerative. Supplier selection is a MCDM problem involving qualitative and quantitative factors influencing the company's key strategies. Specifically, Fuzzy logic is developed from fuzzy set theory to perform the argument approximately instead of an exact argument according to classical predicate logic. Fuzzy logic can be considered the applied side of fuzzy set theory to dealing with real-world values for complex problems. Analytic Network Process (ANP) is an evolved form of Hierarchical Analysis Process (AHP) used for multi-criteria decision making, in which the structure of ANP is a network structure that is, there is an interaction between factors. ANP is applied in many industries and fields such as non-essential consumer goods, manufacturing and manufacturing of electronic components, mobile devices, etc. From which the analysis process in the ANP network can make multi-criteria decisions based on information. The integration of fuzzy logic and ANP can give a better idea than other models like AHP, Fuzzy, and ANP separately. DEA is a linear programming method for measuring and ranking the effectiveness of decisions. On the plus side, DEA, needless to explicitly specify a math form for the production function, proven helpful in discovering relationships that are still hidden for other methods, capable of handling multiple inputs and output, can be used with any input-output measurement, inefficiencies can be analysed and quantified for every unit evaluated. Also, disadvantages of this approach; as a result, is very sensitive to the choice of inputs and outputs, the number of efficient firms at the frontier tends to increase according to the number of inputs and output variables. Therefore, the combination of Fuzzy Analytic Network Process (FANP) and DEA methods that make it easier to look at green issues under uncertain environmental conditions. It also provides a helpful guide for selecting suppliers for RE projects.

2. Literature Review

Many research has frequently applied MCDM methods to select suppliers in the energy sector and various industries nowadays. In many cases, the utilized fuzzy sets theory to establish multi-criteria decisions in an imprecise environment where opinions and judgments are ambiguous is necessary, affecting suppliers' selection criteria or decision support in RE projects (Petrović et al., 2019).

2.1. Multi-Criteria Decision Making

MCDM methods are frequently applied to determine the appropriate criteria in the decision-making process of choosing a supplier of equipment and building materials for RE plants or many other fields, such as (Peng et al., 2020) presented a combined MCDM model based on image fuzzy exponential entropy to determine the weights of suitable supplier selection criteria and an extended Vlse Kriterijumska Optimizacija Kompromiscija Resenje (VIKOR) method to rank sustainable suppliers in the supply chain. (Shojaei et al., 2020) proposes a green supplier selection approach in MCDM construction projects based on Technique for Order Performance by Similarity to Ideal Solution (TOPSIS) and Weighted Aggregated Sum Product Assessment (WASPAS) methods. The research results have suggested the application of raw theory to analyze the behavior of variables in the decision-making process through the criteria of environment and social responsibility. (Pérez-Velazquez et al., 2020) incorporated the diffusion technique with the MCDM model and VIKOR to support supplier selection decision making. The paper shows that the availability of ambiguous data is directly related to supplier ranking position. (Keshavarz Ghorabaee et al., 2021) provided a Multi-Attribute Decision Making (MADM) approach to supplier evaluation and selection based on the two most commonly used methods by researchers, the AHP and TOPSIS methods. (Elleuch et al., 2020) The proposed Multi-Criteria Group Decision Making Method (MCGDM) and TOPSIS method in the fuzzy environment to evaluate sustainable energy sources based on technical, economic, social, political, and environmental issues. The article shows the process of ranking energy sources collected by Fuzzy MCGDM way and using the fuzzy Delphi method to estimate the priority of the criteria. (Çolak et al., 2017) combined the fuzzy MCDM model with fuzzy sets based AHP method and fuzzy TOPSIS method to propose alternative solutions for RE sources in Turkey.

2.2. Methods of applying supplier selection and ranking

(Kumar et al., 2017) applied MCDM modeling techniques to assist decision-makers in making criteria-based decisions in the RE sector. (Wang et al., 2018) proposed a fuzzy MCDM model approach through a fuzzy Hierarchical Analysis Process (FAHP) and DEA methods to solve the problem of selection a power dashboard supplier solar energy for photovoltaic system design in Taiwan under fuzzy environment conditions. Samut & Aktan (2021) presented a methodology for selecting the best supplier for a wind turbine construction project based on qualitative and quantitative factors. The model utilized the FANP method to calculate the weighting of the determining criteria in combination with DEA to evaluate supplier activities. (Hamal et al., 2018) through applying the FANP, has approached considering the unclear judgments of decision-makers' selection investment projects on renewable energy. The objective of this research was to provide a realistic evaluation of energy sources, review, and assist in evaluating investor decisions in fuzzy environment conditions. (Wang et al., 2019) introduced a fuzzy MCDM model to solve wind turbine supplier selection in Vietnam. The model is based on the FANP method to determine the weights involved to the criteria combining the TOPSIS method the ranking of the potential suppliers. (Tavassoli et al., 2020) applied the stochastic-fuzzy DEA (SFDEA) model to evaluate supplier sustainability under undefined, random, and fuzzy conditions. Kaya Samut (2021) proposed the FANP model integrated with the Modified Interactional Group Process (MIGP) model to select the turbine supplier in the most optimal wind power plant project. (Wang et al., 2018) presented the MCGDM model to solve supplier selection problems based on applying the FANP method to calculate the weights of involved criteria combined with the DEA method the ranking of the potential suppliers

This research aims to develop the MCDM model to assist in selecting equipment suppliers to build factory suitable RE to meet Vietnam's energy production sustainable socio-economic development in the context of fuzzy environment conditions. In the proposed model, the FUZZY FANP and DEA methods are used to determine the weights of the criteria and the ranking of the potential suppliers.

3. Methodology

This study was carried out through three steps. The first step decision-makers consult relevant literature, interview managers and experts in the renewable energy field to determine the appropriate criteria and sub-criteria in the selection process supplier. In the second step, the collected data will be processed by the FANP method to determine the weights of the above criteria. After checking the results and reviewing the consistency of the met criteria, these weights will be used for the input and output of the DEA model to evaluate and rank potential suppliers.

3.1. Step 1: Identifying related criteria and sub – criteria

In this case, the supplier selection and evaluation process are determined based on four main criteria and 12 sub-criteria through relevant documents and shown in Figure 3. After identifying the related criteria, potential suppliers are selected to match these criteria.

3.2. Step 2: Applying the FANP method

The decision-making problems involving many quantitative and qualitative criteria in selecting the right supplier are highly complex. Therefore, this study has applied the model combining fuzzy set theory into ANP (supermatrix), the best tool to solve this problem (Sarkis, 1999).

After the managers, experts, and experts in supply chain management completed the questionnaire related to the pairwise comparison matrix to compare the factors side by side. The collected data is used to calculate the weights of supplier metrics as well as the respective priorities of the decision-making units (DMUs) using Super Decision software.

The inconsistency rate was then calculated. If the inconsistency rate is smaller than 0.1, the process can move on to the next step.

3.3. Step 3: Applying the DEA model

The results of the FANP analysis in step 2 are applied in step 3 of our analysis utilizing the DEA analysis.

Figure 1. DEA Model

Each supplier (DMU) has two inputs and two outputs (Figure 1). The results of the FANP model for the ranking of various suppliers on qualitative attributes are utilized in the output 'Qualitative Benefits' of the DEA model. In this case, inputs are considered factors that the organization would consider as an improvement if they were decreased in value, while factors organizations would consider as improvements if they were increased in value are considered outputs (Sarkis, 1999).

The dataset of the DEA model is presented in Table 3. The data set has been normalized to reduce the scaling errors associating with mathematical programming software. The efficiency scores of the ratio-based DEA models are not affected by this normalization process.

Each qualitative variable of each supplier is rated on a 5-point Likert scale. The score of each criterion is based on the opinion of executives within the organization. 5-point scales are common for evaluating in terms of qualitative data, and are often accompanied by interpretations such as: 1 = very bad, 2 = bad, 3 = medium, 4 = good, 5 = very good, which are easily understood by decision-makers.

To complete the analysis, the data set is executed for DEA models using software programs developed in GAMS for this purpose.

4. Case study

In order to demonstrate the model feasibility, the proposed model is applied to a case study where the decision makers must identify an optimal supplier for a solar energy project among 10 potentials candidates. Based on existing literatures and expert review, a list of qualitative criteria is identified (Table. 1):

Table 1. Qualitative benefits

Qualitative Benefit	**Code**
Supply chain coordination improvement	A1
Supplier communication	A2
Upside supply chain flexibility	A3
Downside supply chain flexibility	A4
Internal approvement	A5

After the experts have finished scoring the relative importance of the qualitative benefits, FANP model is applied to calculate the priorities of the DMUs according to the qualitative benefits:

Table 2. DMUs' priorities

DMU	Fuzzy Weights			BNP	Normaliza-tion
DMU 1	0.062	0.114	0.208	0.128	0.114
DMU 2	0.055	0.100	0.183	0.113	0.100
DMU 3	0.071	0.132	0.235	0.146	0.130
DMU 4	0.057	0.105	0.191	0.118	0.105
DMU 5	0.073	0.132	0.237	0.147	0.131
DMU 6	0.041	0.072	0.134	0.082	0.073
DMU 7	0.041	0.073	0.137	0.084	0.075
DMU 8	0.050	0.091	0.163	0.101	0.090
DMU 9	0.054	0.097	0.172	0.108	0.096
DMU 10	0.046	0.084	0.156	0.095	0.085

Finally, the proposed DEA model is applied to calculate the efficiency scores of the DMUs, based on the output of the FANP model. The input data of the DEA model is shown in Table. 3:

Table 3. DEA input data

DMU	Input(s)		Output(s)	
	Unit Price	Delivery time	Quality	Qualitative Benefits
DMU 1	896	23	9	0.1138
DMU 2	790	32	9	0.1004
DMU 3	890	25	8	0.1301
DMU 4	981	24	7	0.1049
DMU 5	900	22	9	0.1314
DMU 6	897	24	8	0.0732
DMU 7	880	23	9	0.0748
DMU 8	990	19	9	0.0903
DMU 9	970	20	9	0.0961
DMU 10	908	21	7	0.0850

The final efficiency scores results are shown in Table. 4:

Table 4. Final efficiency scores of the DMUs

DEA	DEA Model(s)			
	CCR-I	CCR-O	BCC-I	SBM-O
DMU 1	0.9917	0.9917	0.9917	0.9826
DMU 2	1	1	1	1
DMU 3	1	1	1	1
DMU 4	0.7325	0.7325	0.9149	0.7322
DMU 5	1	1	1	1
DMU 6	0.8686	0.8686	0.9762	0.8424
DMU 7	1	1	1	1
DMU 8	1	1	1	1
DMU 9	0.9936	0.9936	0.9936	0.9917
DMU 10	0.7889	0.7889	1	0.7837

From the results, it can be concluded that the optimal suppliers are DMU 2, DMU 3, DMU 5, DMU 7, and DMU 8.

5. Conclusion

While supplier selections are regarded as complex decision-making problem, as they usually involve many quantitative and qualitative criteria. In the field of renewable energy development, there are few tools that can be applied to support such problems. This study aims to utilize the combination of Fuzzy Analytic Network Process (FANP) and DEA methods to create a comprehensive approach to the problem under uncertain decision-making environment. By applying the proposed model to a real-world problem, the feasibility of the model is proved. The model can be extended and applied to similar supplier selection problem in different industries under different sets of qualitative criteria. In future studies, the model performance can be compared with other popular MCDM models such as FAHP-TOPSIS or FANP-DEMATEL.

Reference

Çolak, M., and Kaya, İ. (2017). Prioritization of renewable energy alternatives by using an integrated fuzzy MCDM model: A real case application for Turkey. *Renewable And Sustainable Energy Reviews*, 80, 840-853.

Elleuch, M., Mallek, M., Frikha, A., Hachicha, W., Aljuaid, A., & Andejany, M. (2021). Solving a Multiple User Energy Source Selection Problem Using a Fuzzy Multi-Criteria Group Decision-Making Approach. *Energies*, 14(14), 4313.

Hamal, S., Senvar, O., & Vayvay, O. (2018). Selection of optimal renewable energy investment project via fuzzy ANP. *Pressacademia*, 5(2), 224-233.

Kaya Samut, P. (2017). Integrated FANP-f-MIGP model for supplier selection in the renewable energy sector. *Journal Of Business Economics And Management*, *18*(3), 427-450.

Keshavarz Ghorabaee, M., Amiri, M., Zavadskas, E., & Antucheviciene, J. (2017). Supplier evaluation and selection in fuzzy environments: a review of MADM approaches. *Economic Research-Ekonomska Istraživanja*, *30*(1), 1073-1118.

Kumar, A., Sah, B., Singh, A., Deng, Y., He, X., Kumar, P., & Bansal, R. (2017). A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renewable And Sustainable Energy Reviews*, 69, 596-609.

Petrović, G., Mihajlović, J., Ćojbašić, Ž., Madić, M., & Marinković, D. (2019). Comparison of three fuzzy mcdm methods for solving the supplier selection problem. *Facta Universitatis, Series: Mechanical Engineering*, 17(3), 455.

Peng, J., Tian, C., Zhang, W., Zhang, S., & Wang, J. (2020). An integrated multi-criteria decision-making framework for sustainable supplier selection under picture fuzzy environment. *Technological And Economic Development Of Economy*, 26(3), 573-598.

Pérez-Velázquez, A., Oro-Carralero, L., & Moya-Rodríguez, J. (2020). Supplier Selection for Photovoltaic Module Installation Utilizing Fuzzy Inference and the VIKOR Method: A Green Approach. *Sustainability*, 12(6), 2242.

Samut, P., & Aktan, H. (2021). *Integrating qualitative and quantitative factors in supplier selection and performance evaluation.*

Shojaei, P., & bolvardizadeh, A. (2020). Rough MCDM model for green supplier selection in Iran: a case of university construction project. *Built Environment Project And Asset Management*, 10(3), 437-452.

Tavassoli, M., Saen, R., & Zanjirani, D. (2020). Assessing sustainability of suppliers: A novel stochastic-fuzzy DEA model. *Sustainable Production And Consumption*, 21, 78-91.

Wang, C., Nguyen, V., Duong, D., & Do, H. (2018). A Hybrid Fuzzy Analytic Network Process (FANP) and Data Envelopment Analysis (DEA) Approach for Supplier Evaluation and Selection in the Rice Supply Chain. *Symmetry*, 10(6), 221.

Wang, C., Yang, C., & Cheng, H. (2019). Fuzzy Multi-Criteria Decision-Making Model for Supplier Evaluation and Selection in a Wind Power Plant Project. *Mathematics*, 7(5), 417.

Wang, T., & Tsai, S. (2018). Solar Panel Supplier Selection for the Photovoltaic System Design by Using Fuzzy Multi-Criteria Decision Making (MCDM) Approaches. *Energies*, 11(8), 1989.

LOCATION SELECTION PROCESS FOR GEOTHERMAL ENERGY PROJECTS IN VIETNAM: AN MCDM APPROACH

Viet Tinh Nguyen [1,*], Phi Truong Cao [2], Thi Tuyet Mai Truong [2] and Thi Ngoc Vu [2]

[1]*Faculty of Commerce, Van Lang University, Vietnam, tinh.nv@vlu.edu.vn*

[1]*Faculty of Commerce, Van Lang University*

Abstract. Geothermal is a renewable energy with great economic and environmental potential as reduce and emit greenhouse gas. Geothermal energy has been produced and utilized since early 20th century for the purposes of heating, drying agricultural products, relaxing baths, ect . current production technologies have achieved a degree of maturity, allowing the development of large geothermal projects in developed countries such as the USA, China, and Japan. This energy source is being used in 70 countries around the world. Vietnam has great potential for renewable energy production such as solar, wind, biomass, and geothermal energy. The research and development of geothermal energy is growing rapidly in scale and efficiency, project can also bring new opportunities to Vietnam in technological development, job creation, and most importantly, a potential for a green economy. The effectiveness of geothermal projects depends heavily on the location selection decision-making process due to the nature of geothermal energy. Furthermore, the decision-makers also must consider multiple economic, social, and environment criteria, in addition to the technological requirements of the project. Currently, Vietnam has significant geothermal potential and can develop thermal power plants but there is not much research on locating thermal power plants using MCDM multi-criteria decision model. Compared to other energy sources, the impact of geothermal power plants on the environment will depend heavily on their location. This research aims to develop a fuzzy MCDM approach to the location selection problem of geothermal energy projects in Vietnam utilizing Fuzzy Analysis Network Process (FANP) and The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) methods to be able to determine the optimal location, or complex problems when placing the factory at that location so that the most useful decisions and solutions can be made. The proposed model is then applied to a real-world case study in Vietnam.

Keywords: Renewable energy; Geothermal energy, Supply chain management; MCDM; Location selection; FANP; TOPSIS.

1. Introduction

Geothermal energy is considered a form of renewable energy because it is available and environmental friendly. According to the research of Quang Khanh Do et al. (2018), Vietnam is considered to have average geothermal potential compared to the world of up to 400 MW capacity. The geothermal potential in Vietnam estimate more than 300 geothermal energy sources in Vietnam territory with the temperatures from 30 °C to 148 °C . However, many geothermal energy projects in Vietnam mostly remain at the researching stage and are still under-development due to many reasons. In which the selection of the location of a geothermal plant is a laborious task and critical concern because the effectiveness of geothermal projects depends heavily on the location selection decision-making process due to the nature of geothermal energy.

Currently, there is a little research on geothermal power plant location determination uses multi-criteria decision model (MCDM) and there is still no method to evaluate and select the optimal location for these geothermal power plants geothermal power plants while considering environmental issues. Further, location selection is complicated, in that decision-makers must have broad perspectives concerning qualitative and quantitative criteria. In this research, the authors will use the criteria decision-making model (MCDM) by combining two methods approach to the location selection problem of geothermal energy projects in Vietnam utilizing Fuzzy Analysis Network Process (FANP) and Ideal Solution Similar Prioritization (TOPSIS) technique to find the best location to build a geothermal plant in Viet Nam.

Firstly, the authors using FANP for defining the weight of each potential location, because FANP model is the most effective tool for solving complex problems of decision-making involving different characteristics. However , in the FANP model the value of the criteria is provided by the experts so are that the involves subjectivity. Therefore, the authors proposed the TOPSIS model to rank potential sites in order to solve the disadvantage of the FANP method.

As mentioned ,the purpose of this research is to propose a suitable MCDM model with combination of two FANP-TOPSIS approach presented to selection the best location for a geothermal plant in Vietnam.

2. Literature Review

Currently, the multi-criteria decision-making model is widely applied in the field of renewable energy. Location selection for renewable energy plants is one of the areas where researchers have applied the MCDM model. Research has frequently applied MCDM methods to select suppliers in the energy sector and various industries nowadays. In many cases, the utilized fuzzy sets theory to establish multi-criteria decisions in an imprecise environment where opinions and judgments are ambiguous is necessary, affecting suppliers' selection criteria or decision support in RE projects (Petrović et al., 2019).

2.1. The multi-criteria decision-making (MCDM) model in renewable energy

The multi-criteria decision-making model (MCDM) has been widely applied in the field of renewable energy. Many researchers have applied this model. (Indre Siksnelyte et al. 2020) for a review of Household Renewable Energy Technologies proposed Multi-Criteria Decision Making (MCDM). (Arash Sadeghi et al., 2012) propose a fuzzy multi-criteria decision-making method (FMCDM) to evaluate 4 alternative renewable energy types including solar, geothermal, electric power and wind energy in Yazd Province, Iran. (Sarmad Ishfaq et al., 2018) used the MCDM approach to select the optimal renewable energy source for the energy sector in Pakistan. (Gulcin et al., 2017) evaluated renewable energy sources in Turkey using an integrated MCDM approach with linguistic interval preference relationships. (Yunna Wu et al., 2018) based on cumulative prospect theory to evaluate renewable power sources using fuzzy MCDM. (Yazdani-Chamzini et al., 2013) select the optimal renewable energy by making multi-criteria decision.

2.2. Approach of the ranking and location selection

Ugo, (2015) proposed a multi-criteria decision-making model using Fuzzy (TOPSIS) for Location Selection in Niger Delta. (Chia Nan Wang et al., 2018) proposed multi-criteria decision making, by combining three methods, including fuzzy analysis hierarchical process (FAHP), data packaging analysis (DEA) and ordering techniques are prioritized by similarities with the ideal solution (TOPSIS) to find the best location for the construction of solar power plants in Vietnam. (Chia Nan Wang et al., October 2018) to prepare for the selection of nuclear power plant locations in Vietnam, two main methods were used including fuzzy analytical network (FANP) process and technique of prioritization by the similarity of an ideal solution (TOPSIS). (Chia Nan Wang et al., 2020) consider the decentralized integrated model (DEA) and (FANP) for the optimal geographical location of solar power plants in the Mekong Delta Region, Vietnam. (Lee AHI et al., 2017) using a comprehensive multi-criteria decision-making model, combining interpretive structure model (ISM), fuzzy analytical network process (FANP) and (VIKOR) is proposed to choose the most suitable photovoltaic plant location. (Asad Asadzadeh et al., 2014) to assess the location selection of New Towns of the Tehran Metropolitan Region used the TOPSIS model.

This research will provide background documents to support the development of the MCDM model and the proposed FANP-TOPSIS combination method to select the best location to build a geothermal plant in Vietnam. At the end of this research., the main discussions and contributions of the study will be presented.

3. Methodology

3.1. Fuzzy analytic network process (FANP) method

Due to Fuzzy analytic network process (FANP) relative simplicity in comparison with Fuzzy analytical hierarchy process (FAHP), FANP is frequently used as an alternative to FAHP in calculating priority weights from fuzzy comparison matrices (Wang, 2018).

Let $X = \{x_1, x_2, x_{3,} \dots . x_n\}$ be an object set and $O = \{o_1, o_2, o_{3,} \dots . o_n\}$ be a set of goals. The FANP process takes each object from set X then conduct an extended analysis of each goal (o_i) of the object. As such, the extent analysis values of each object (x_i), v, can be shown as followed:

$$V_{o_i}^1, V_{o_i}^2, \dots, V_{o_i}^m, \quad i = 1,2, \dots, n \tag{2}$$

Where $V_{o_i}^j (j = 1,2, \dots., m)$ are the TFNs.

The extended analysis process can be shown as followed:

Step 1: Determine the fuzzy synthetic extent value of the i^{th} object as:

$$S_i = \sum_{j=1}^{m} V_{o_i}^j \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} V_{o_i}^j \right]^{-1} \tag{3}$$

With the fuzzy addition operation of *m* extent analysis values of the object matrix $(\sum_{j=1}^{m} V_{o_i}^j)$ determined by:

$$\sum_{j=1}^{m} V_{o_i}^j = \left(\sum_{j=1}^{m} r_j, \sum_{j=1}^{m} p_j, \sum_{j=1}^{m} q_j, \right). \tag{4}$$

The fuzzy additional operation of $V_{a_i}^j (j = 1,2, \dots., m)$ values $([\sum_{i=1}^{n} \sum_{j=1}^{m} V_{o_i}^j]^{-1})$ are calculated by:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} V_{o_i}^j = \left(\sum_{j=1}^{n} r_j, \sum_{j=1}^{n} p_j, \sum_{j=1}^{n} q_j, \right). \tag{5}$$

Then, the inversion of the vector in (5) is determined by:

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{m} V_{o_i}^j \right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} r_i}, \frac{1}{\sum_{i=1}^{n} p_i}, \frac{1}{\sum_{i=1}^{n} q_i} \right). \tag{6}$$

Step 2: The degree of possibility of $V_2 = (r_2, p_2, q_2) \geq V_1 = (r_1, p_1, q_1)$ is calculated as:

$$\mathrm{P}(V_1 \geq V_2) = sup_{y \geq x} \left[min(\mu_{V_1}(x),), (\mu_{V_2}(y)) \right] \tag{7}$$

Which also be shown as:

$$\mathrm{P}(V_1 \geq V_2) = hgt(V_1 \frown V_2) = \mu_{V_2}(d) = \begin{cases} 1 & if p_2 \geq p_1 \\ 0 & if r_1 \geq r_2 \\ \frac{r_1 - q_2}{(p_2 - q_2) - (p_1 - q_1)} & otherwise \end{cases} \tag{8}$$

where d is the ordinate of the highest intersection point D between μ_{v_1} and μ_{v_2}. In order to be able to compare V_1 and V_2, we need to calculate the possibility of $(V_1 \geq V_2)$ and $(V_2 \geq V_1)$.

Step 3: Calculate the degree of the possibility that a convex fuzzy number is greater than c convex fuzzy number, with $V_i (i = 1,2, \dots, c)$ as:

$$\mathrm{P}(V \geq V_1, V_2, \dots, V_k) = P[(V \geq V_1) \; and \; (V \geq V_2) \;] \tag{9}$$

and,

$$(V \geq V_c) = \min \mathrm{P}(\mathrm{V} \geq V_i), i = 1,2, \dots, c$$

Assuming that $d'(B_i) = minP(S_i \geq S_c)$, for $c = 1,2, \dots n$ and $c\#i$, determine the weighted vector as:

$$W' = (d'(B_1), d'(B_2), \dots d'(B_n))^H, \tag{10}$$

Where A_i are n elements.

Step 4: Calculate the normalized weighted vector:

$$\mathrm{d}(B_i) = \frac{d'(B_i)}{\sum_{i=1}^{n} d'(B_1)} \tag{11}$$

$$\mathrm{W} = (d(B_1), d(B_2), \dots., d(B_n))^H \tag{12}$$

3.2. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

Hwang et al. (1981) first proposed the TOPSIS method. A classic TOPSIS process includes 5 steps as followed:

Step 1: Establish the normalized decision matrix.

$$g_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \quad (13)$$

where x_{ij} is the original score of decision matrix and g_{ij} is the normalized score of the decision matrix.

Step 2: Determine the normalized weight matrix:

$$o_{ij} = h_j g_{ij} \quad (14)$$

here h_j represents the weight of the j^{th} criterion.

Step 3: The Positive ideal solution (PIS) matrix and Negative ideal solution (NIS) matrix are defined as:

$$O^+ = o_1^+, o_2^+, \dots, o_n^+$$
$$O^- = o_1^-, o_2^-, \dots, o_n^-; \quad (15)$$

Step 4: Calculate the performance gap between values of each option using the PIS and NIS matrices.

The distance to PIS of each option is calculated as:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} (o_i^+ - o_{ij})^2} \;;\; i = 1, 2, \dots, m \quad (16)$$

The distance to NIS of each option is calculated as:

$$D_i^- = \sqrt{\sum_{j=1}^{m} (o_{ij} - o_i^-)^2} \;;\; i = 1, 2, \dots, m \quad (17)$$

with D_i^+ as the distance to the PIS and D_i^- as the distance to the NIS for the i^{th} option.

Step 5: The preference value (V_i) of each option is calculated as:

$$V_i = \frac{D_i^-}{D_i^- + D_i^+} \qquad i = 1, 2, \dots, m \quad (18)$$

Finally, the V_i values are used to evaluate and rank the potential options.

4. Case Study

Vietnam is a welcoming country with a number of renewable energy projects especially geothermal energy. However, geothermal energy requires a strict number of criteria in order to evaluate each location thoroughly. A total of ten criteria have considered according to Uliasz-Misiak et al (2021) as the authors considered when determining geothermal selection sites. The ten criteria are listed in Tab. 1,

Table 15: List of Criteria in determining geothermal selection site

No.	Criteria
1	Geological structure
2	Earth's thermal flux
3	Rocks temperatures
4	Hydraulic conductivity of aquifers and ground water
5	Geothermal water availability
6	Depth of drilling
7	Water mineralization
8	Energy price
9	Heat price
10	Location of potential wells

The study observed a total of 9 suitable locations all around Vietnam which will denoted as alternatives so that the methodology can be applied in order to calculate the best suitable location.

5. Results and Discussion

The authors will conduct the normalization of criteria based on the experts opinion and literature review. The FANP process will be applied initially in order to determine the overall weight of each criteria. Tab. 2 first shows the fuzzy geometric mean for each criteria.

Table 2: Fuzzy Geometric mean of each criteria

Criteria	Fuzzy Geometric Mean of Each Row		
C1	0.8112	1.1710	1.6571
C2	0.7257	0.9915	1.3623
C3	0.9931	1.3979	1.8895
C4	0.8542	1.1791	1.5547
C5	1.0043	1.3798	1.8535
C6	0.5395	0.7095	0.9864
C7	0.5830	0.7759	1.0729
C8	0.6856	0.9310	1.2536
C9	0.7222	0.9579	1.2485
C10	0.5808	0.7715	1.0828

The criteria are then applied with FANP methodology in order to determine fuzzy weights as shown in Tab. 3.

Table 3: Fuzzy weights of each criteria

Criteria	Fuzzy Weights		
C1	0.0581	0.1141	0.2210
C2	0.0520	0.0966	0.1817
C3	0.0711	0.1362	0.2519
C4	0.0612	0.1149	0.2073
C5	0.0719	0.1344	0.2471
C6	0.0386	0.0691	0.1315
C7	0.0418	0.0756	0.1431
C8	0.0491	0.0907	0.1672
C9	0.0517	0.0933	0.1665
C10	0.0416	0.0752	0.1444

The criteria is then normalized in order to determine the final weights of each criteria (shown in Tab. 4) that will be used to determine the final alternative best suited.

Table 4: Final normalized weight for 10 criteria

Criteria	Normalization
C1	0.1157
C2	0.0972
C3	0.1351
C4	0.1128
C5	0.1334
C6	0.0704
C7	0.0766
C8	0.0903
C9	0.0917
C10	0.0768

The alternatives are now first normalized as the first step of TOPSIS shown in Tab. 5 and Tab. 6.

Table 5: Normalized matrix for Criteria 1-5 for 9 alternatives

	C1	C2	C3	C4	C5
A1	0.4456	0.1768	0.1504	0.2817	0.3546
A2	0.3961	0.2946	0.2005	0.1690	0.3152
A3	0.3961	0.4125	0.4511	0.4507	0.3546
A4	0.2475	0.1768	0.4010	0.1127	0.3152
A5	0.1485	0.5303	0.4010	0.5071	0.3152
A6	0.3961	0.2946	0.2506	0.2254	0.3152
A7	0.0990	0.4125	0.4511	0.4507	0.3546
A8	0.4456	0.2946	0.3509	0.2254	0.3546
A9	0.1980	0.2357	0.1504	0.3381	0.3152

Table 6: Normalized matrix for Criteria 6-10 for 9 alternatives

	C6	C7	C8	C9	C10
A1	0.3690	0.3913	0.4711	0.1558	0.1442
A2	0.3690	0.2174	0.2617	0.2077	0.3845
A3	0.3690	0.3913	0.2617	0.3634	0.1922
A4	0.3229	0.3913	0.4187	0.4673	0.3364
A5	0.2306	0.2174	0.2094	0.4153	0.4325
A6	0.3229	0.3913	0.2617	0.3115	0.1922
A7	0.4151	0.2174	0.3664	0.2077	0.4325
A8	0.2306	0.3043	0.2094	0.3115	0.2883
A9	0.3229	0.3913	0.4187	0.4153	0.4325

The alternatives are then multiplied accordingly based on the weights of the criteria and is shown in Tab. 7 and 8.

Table 7: Normalized weighted matrix for Criteria 1-5 for 9 alternatives

	C1	C2	C3	C4	C5
A1	0.0515	0.0172	0.0203	0.0318	0.0473
A2	0.0458	0.0286	0.0271	0.0191	0.0421
A3	0.0458	0.0401	0.0610	0.0508	0.0473
A4	0.0286	0.0172	0.0542	0.0127	0.0421
A5	0.0172	0.0515	0.0542	0.0572	0.0421
A6	0.0458	0.0286	0.0339	0.0254	0.0421
A7	0.0115	0.0401	0.0610	0.0508	0.0473
A8	0.0515	0.0286	0.0474	0.0254	0.0473
A9	0.0229	0.0229	0.0203	0.0381	0.0421

Table 8: Normalized weighted matrix for Criteria 6-10 for 9 alternatives

	C6	C7	C8	C9	C10
A1	0.0260	0.0300	0.0425	0.0143	0.0111
A2	0.0260	0.0167	0.0236	0.0190	0.0295
A3	0.0260	0.0300	0.0236	0.0333	0.0148
A4	0.0227	0.0300	0.0378	0.0428	0.0258
A5	0.0162	0.0167	0.0189	0.0381	0.0332
A6	0.0227	0.0300	0.0236	0.0286	0.0148
A7	0.0292	0.0167	0.0331	0.0190	0.0332
A8	0.0162	0.0233	0.0189	0.0286	0.0222
A9	0.0227	0.0300	0.0378	0.0381	0.0332

The final ranking for each alternative after considering all the criteria is then shown in Tab. 9.

Table 9: Final ranking of each alternative after considering all the criteria

Alternative	Si+	Si-	Ci	Rank
A1	0.0692	0.0532	0.4344	8
A2	0.0657	0.0433	0.3976	9
A3	**0.0317**	**0.0742**	**0.7008**	**1**
A4	0.0623	0.0553	0.4702	5
A5	0.0467	0.0735	0.6113	2
A6	0.0572	0.0460	0.4457	7
A7	0.0511	0.0674	0.5687	3
A8	0.0531	0.0551	0.5092	4
A9	0.0614	0.0494	0.4461	6

From Tab. 9, option A3 gives the best alternatives for all of the possible criteria and alternative by having the highest overall criteria index. From there the decision maker can decide suitably the best location of the new renewable energy project.

6. Conclusion

From the study, the FANP and TOPSIS methods have been used successfully in order to bring the best suitable results that would assist decision-makers specifically project managers in deciding the best possible alternative when deploying renewable energy projects. Further studies of other MCDM methods can be applied in order to expand the study for better assistance.

Reference

Asadzadeh, A., Sikder, S., Mahmoudi, F. and Kötter, T., 2014. Assessing Site Selection of New Towns Using TOPSIS Method under Entropy Logic: A Case study: New Towns of Tehran Metropolitan Region (TMR). *Environmental Management and Sustainable Development*, 3(1), p.123.

Büyüközkan, G. and Güleryüz, S., 2017. Evaluation of Renewable Energy Resources in Turkey using an integrated MCDM approach with linguistic interval fuzzy preference relations. *Energy*, 123, pp.149-163.

Destiny Ugo, P. (2015). A Multi-Criteria Decision Making for Location Selection in the Niger Delta Using Fuzzy TOPSIS Approach. *International Journal of Management and Business Research*, 5(3), 215-224.

Do, Q., Doan, T., Hoang, T., Nguyen, T. and Tran, T., 2018. Potential Application of Slimhole Drilling Technology to Geothermal Wells in Vietnam. *Lecture Notes in Civil Engineering*, pp.239-244.

Hwang, C.L.; Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. New York: Springer-Verlag.

Lee, A., Kang, H. and Liou, Y., 2017. A Hybrid Multiple-Criteria Decision-Making Approach for Photovoltaic Solar Plant Location Selection. *Sustainability*, 9(2), p.184.

Sadeghi, A., Larimian, T. and Molabashi, A., 2012. Evaluation of Renewable Energy Sources for Generating Electricity in Province of Yazd: A Fuzzy Mcdm Approach. *Procedia - Social and Behavioral Sciences*, 62, pp.1095-1099.

Siksnelyte-Butkiene, I., Zavadskas, E. and Streimikiene, D., 2020. Multi-Criteria Decision-Making (MCDM) for the Assessment of Renewable Energy Technologies in a Household: A Review. *Energies*, 13(5), p.1164.

Uliasz-Misiak, B., Lewandowska-Śmierzchalska, J., & Matuła, R. (2021). Criteria for selecting sites for integrated CO2 storage and geothermal energy recovery. Journal Of Cleaner Production, 285, 124822. https://doi.org/10.1016/j.jclepro.2020.124822

Wang, C. N., Nguyen, V. T., Duong, D. H., & Thai, H. T. N. (2018). A hybrid fuzzy analysis network process (FANP) and the technique for order of preference by similarity to ideal solution (TOPSIS) approaches for solid waste to energy plant location selection in Vietnam. Applied Sciences, 8(7), 1100.

Wang, C., Nguyen, V., Thai, H. and Duong, D., 2018. Multi-Criteria Decision Making (MCDM) Approaches for Solar Power Plant Location Selection in Viet Nam. *Energies*, 11(6), p.1504.

Wang, C., Su, C. and Nguyen, V., 2018. Nuclear Power Plant Location Selection in Vietnam under Fuzzy Environment Conditions. *Symmetry*, 10(11), p.548.

Wang, C., Viet, V., Ho, T., Nguyen, V. and Husain, S., 2020. Optimal Site Selection for a Solar Power Plant in the Mekong Delta Region of Vietnam. *Energies*, 13(16), p.4066.

Wu, Y., Xu, C. and Zhang, T., 2018. Evaluation of renewable power sources using a fuzzy MCDM based on cumulative prospect theory: A case in China. *Energy*, 147, pp.1227-1239.

Yazdani-Chamzini, A., Fouladgar, M., Zavadskas, E. and Moini, S., 2013. SELECTING THE OPTIMAL RENEWABLE ENERGY USING MULTI CRITERIA DECISION MAKING. *Journal of Business Economics and Management*, 14(5), pp.957-978.

REAL ESTATE BUSINESS IN VIETNAM AND GREEN MARKETING

Nguyen Minh Ngoc[1] and Nguyen Hoang Tien[2]

[1] *University of Finance – Marketing, Ho Chi Minh City, nmngoc@ufm.edu.vn, 0912359472*

[2] *University of Food Industry, Ho Chi Minh City, vietnameu@gmail.com*

Abstract: Green Marketing has never been more important and powerful as world trade has moved towards being environmentally friendly in the past few decades. Environmental pollution has significantly affected the economic development of Vietnam in the period of international integration, especially the post-Covid-19 period. The boom of the real estate market in the past 20 years and further economic liberalization is growing rapidly in emerging markets in line with increasing green marketing's capabilities with a wide range of communication channels, giving new customers more access to the benefits that green marketing offers for a better quality of life, more environmentally friendly and sustainable real estate products. The research will demonstrate how the post-Covid-19 green marketing development strategy in Vietnam will affect and interact in the real estate business.

Keywords: Real estate businesses, development strategy, green marketing.

1. Introduction

According to Cambridge dictionary, Green Marketing refers to the marketing activities of products that are considered good, environmentally friendly, expressed in the strategies of changing product design, packaging process, advertising activities to respond to meet the green needs of the user.

After struggling with an economy heavily damaged by the environment, the COVID-19 pandemic is happening on a global scale. Originated in December 2019 with the first recorded epidemic in the central Chinese city of Wuhan, stemming from a group of people with unexplained pneumonia. Local health officials confirmed that they have had prior contact, mainly with traders who trade and work at the Hainan seafood wholesale market, which is believed to be the first outbreak of the disease. A series of green movements emerged and became a global trend. The birth of many green concepts such as green real estate products, green urban area to promote green consumption in the era of green development, to repel the Covid-19 epidemic.

The Green Marketing trend has emerged and become a direct competitor to traditional marketing because this strategy is aimed at the benefits of the environment, the benefits of a clean, disease-free country, which is also a concern of top of the countries globally. Moreover, in addition to the practical benefits from the post-Covid 19 environment, real estate enterprises now aim to develop sustainable to create a competitive advantage over competitors and in line with many policies promoting Vietnam's post-epidemic environment-friendly economy. Both have the effect of awakening the conscience and protecting the environment after the pandemic outbreak. Especially in the field of real estate, which consumes a lot of land, building materials and energy, green marketing is an important basis for investors to develop towards green products.

2. Theoretical framework

2.1. Green marketing

Almost many people think that Green Marketing just stops at communication activities or advertising about real estate products with a few features related to the environment. A few terms like recycling, reuse, eco-friendly, ozone-friendly are a few things that most consumers associate when it comes to Green Marketing. In fact, Green Marketing does not stop there, it can be applied to materials for construction, home appliances or real estate services.

Therefore, green marketing in real estate involves a lot of investment and development activities and real estate business such as design, construction method, material selection, planning as well as communication. This is not a simple task.

The concept of Green Marketing activity was first proposed by Hennion and Kinnear (1976) when previous works have recommended the importance of providing solutions to be able to overcome these problems negative impact on the environment of marketing activities.

Polonsky (1994) argues that Green Marketing is a combination of a range of activities including product modification, manufacturing process modification, packaging modification, as well as advertising alteration to create and promote any exchange to satisfy the needs and wants of consumers on the basis of minimizing negative impacts on the environment.

Peattie (1995) argues that Green Marketing is a comprehensive governance process, responsible for defining tasks such as predicting, responding to the needs of consumers and society in a sustainable manner. According to this understanding, we confirm that there really exists a relationship between ethics and Green Marketing. Environmental and planning factors are seen as part of the ethical questions that marketers need to analyze and answer. Therefore, Green Marketing should be evaluated as part of social responsibility (Polonsky, M.J., Rosenberger, P.J., 2001).

Ottoman (1993) found that the definitions of Green Marketing during this period were quite limited when they only focused on emphasizing the influence of enterprises on the environment but not towards persuading them change that negative behavior. Ottman believes that Green Marketing needs to focus on developing products that satisfy all consumer needs, including quality, affordability, usability and compatibility, i.e. minimizing negative impact on the natural environment.

Recognizing these limitations, Peattie K. (2001) described Green Marketing as marketing activities aimed at reducing the negative environmental and social impacts of existing products and production systems and at the same time encourage products and services to have less influence.

More specifically, Kotler (2011) sees Green Marketing as the commitment of a business by providing safe and environmentally friendly products and services by using bags. The packaging is recyclable and biodegradable, adopts better pollution prevention methods and uses more energy efficiently. In short, Green Marketing is not a completely separate form of marketing but still has some overlap with other forms of marketing. This has led to a misunderstanding of the nature of Green Marketing (Dean, T.J., Pacheco, D.F., 2014). The difference of Green Marketing lies in its content and attached to the human values behind when businesses choose to use this marketing direction (Peattie K. 1992).

2.2. Green marketing development strategy in real estate business

Today's environmental crisis is the most appropriate time to reassess the 4Ps (Product, Price, Promotion, and Place) of traditional marketing policies. Green marketers need to receive information regarding these new issues and reassess the marketing mix in a greener direction (Peattie K. 1992).

Green real estate product strategy

Green real estate products, also known as eco-friendly products or environmentally conscious products (Tseng, S., Hung, S., 2013), are understood to be products designed to minimize the impact on the environmental factor in the whole its life cycle (Albino, et al., 2009). Such reduction is shown in many aspects such as reducing the use of non-renewable resources or toxic materials and increasing the use of renewable resources (Robert, 1995). Ottman (1998) asserts that green products are more durable and less toxic because they are derived from recyclable materials. A green product strategy often includes activities such as recycling, reuse; or reduce fabrication materials, reduce packaging materials, increase product durability in use and distribution process (Kinoti, 2011).

More specifically, environmentally friendly real estate requires methods of conserving natural resources; low carbon emissions. There are some examples of natural resource conservation in landscape planning such as low construction density, trees and water surface accounting for a large proportion; use insulation to limit energy loss; apply measures to reduce solid waste discharge into the environment; use of solar energy (Tien, N.H., Ngoc, N.M., 2019).

That said, the concern about green products has ceased to be the same as the previous theories, as consumers are not just looking for traditional tangible products. Businesses overestimate the green attributes of traditional products, leading to consumer refusal or even a negative reaction (Davari, A., Strutton, D., 2014). As a result, some companies are creating products that offer environmental value but combine these with functional appeal and emotional element (Dean, T.J., Pacheco, D.F., 2014). Finally, in order to perfect the green product policy, Peattie K. (1992) pointed out that businesses should increase their after-sales support, as green customers tend to see these after-sales services as a criterion to beat price of the product's durability. And green real estate is not out of the way.

Green logistics real estate is one of the current trends that logistics enterprises in countries have been applying for their logistics activities to achieve economic efficiency issues, but still ensure environmental friendliness and achieve social responsibility.

Typically, logistics is seen as the actions of which the objective is to minimize costs and maximize profits. But, for many years, the term logistics was used in conjunction with the "green" by creating "Green Logistics". The term “green logistics real estate” is defined as strategies for designing, building and operating logistics chain management to reduce the environmental and energy impact of goods distribution, focusing on into material handling, waste management, packaging and transportation (Khoa, H.D., Nhung, T.T.B., 2020).

Green marketing activities associated with green logistics real estate business are both beneficiaries and have a positive and effective impact on green logistics businesses. Green marketing costs in the green logistics real estate business tend to be lower than marketing costs for conventional logistics properties. But business efficiency is often higher due to many objective and subjective reasons.

Green real estate pricing strategy

Green prices are defined as the prices set for green products that offset the price sensitivity of consumers, making them willing to pay more for the green attributes of the product (Grove et al., 1996). This excess fee is absolutely necessary when Peattie and Crane (2000) find that green product production requires higher costs than the traditional use of raw materials and production methods, as well as the burden of environmental taxes on the rise. To get green real estate projects or green urban areas, developers have to spend a large initial investment. However, they can dramatically increase product competitiveness and add huge value to customers in the long run (Tien, N.H., Ngoc, N.M., 2019).

Green promotion strategy

Davari and Strutton (2014) argue that promotion is considered as the most important tool of the mixed Green Marketing policy. Green promotion tools are often used to convey messages to encourage customers to become “greener”. In real estate investment and business activities, it is necessary to be flexible in implementing this strategy, and at the same time, it is necessary to achieve the following 3 criteria: (1) the policy directly or implicitly addresses and / or enhances the relationship between products and the biological environment physical; (2) endorsing a green lifestyle with or without a prominent product / service; (3) represent, enhance, or maintain the corporate image associated with environmental responsibility (Banerjee, Gulas, & Iyer, 1995). In the real estate business, the green product promotion strategy requires social interactions that are not only broad but also academic and scientific in depth to achieve the goal (Ngoc, N.M. et al., 2020).

Green distribution strategy

Green distribution refers to management tactics related to the distribution of green goods, from production to consumption and the logistics of recovery (Davari, A., Strutton, D., 2014). In fact, there are very few users trying to find a greener real estate product (Guyader, H. et al., 2017). Therefore, how and where green products are available is the key to a real estate business's survival. Consumers must continuously be exposed to green products throughout the market area. Real estate businesses often use famous, influential people to use their green real estate to assert green values.

Besides, with its distinctive characteristics, selling green products is not the same as selling traditional products. For this reason, Green Marketing staff will manage products from production to point of sale and finally to customers, in order to maximize distribution efficiency (Tomasin, L. et al., 2013). The green real estate distribution channels themselves also need to build their own green brand.

3. Methodology

Authors collect documents with related content, studies different documents and theories by analyzing on the basis of the theory learned in the subject content. After grouping and analyzing the references, the system will be systematized and synthesized into an exercise focusing on the main content of the essay.

To contribute to the research points, the authors conducted in-depth interviews with experts to identify issues that hinder green marketing activities and inadequacies in developing green real estate products in Vietnam.

The authors apply their own experience as both a consumer and a real estate business specialist at universities to come up with solutions for real estate businesses to achieve their goals, be successful with green marketing.

4. Research results and discussion

4.1. Actual situation of implementing green real estate strategy in Vietnam

Like many other countries, Vietnam is currently facing serious environmental pollution, which is a warning bell for businesses to change their sense of production and environmental responsibility.

A project being certified green means that they have met certain verifiable standards. In Vietnam today, there are quite a few green standards in circulation such as: Edge (of the IFC organization of the World Bank); Green Mark (Singapore), LEED (USA)... Each standard has different strengths and weaknesses, but in general, the basic elements are: Sustainable construction site; efficient use of energy and water; using environmentally friendly materials and resources, etc. are basic factors. This means that if building a green building, it must be green from within its inner.

In May 2016, the Ministry of Construction also issued Decision No. 419 on "Action plan of the construction industry on green growth to 2020, vision to 2030". According to this plan, by 2030, 50% of large and medium-sized cities will meet the green city criteria.

Currently, the Ministry of Construction is studying, developing and promulgating specific incentive mechanisms (taxes, fees, etc.) for green buildings that use energy economically and efficiently, and the works are certified as public works green program. At the same time, the Ministry will develop and implement regulations on assessment, certification, labeling and certification of materials, equipment, energy-saving works and green buildings. In addition, the Ministry also encourages strengthening communication activities, raising awareness for stakeholders about economical and efficient use of energy, development of green buildings, implementation of an eco-friendly lifestyle environment, reducing greenhouse gas emissions.

The market has seen the fact that green urban projects bring benefits to many actors. The investor will sell faster, the price will be 4-8% higher. Customers reduce electricity and water costs by 15-20%. In addition, in a recent Nielsen global study on corporate social responsibility, 64% of study participants (from the Asia-Pacific region) said they would be willing to pay more for use sustainable features, green buildings (higher than 50% of 2012).

The number of LEED-certified projects in Vietnam in 2017 was less than 3%. Currently, there are only 150 certified green real estate projects in Vietnam as well as in the process of construction and design. According to statistics of the Green Building Council of other countries, Vietnam is much slower in the number of green buildings as well as in the field of training and awareness.

The real estate businesses in Vietnam themselves, participating in green building construction are also quite modest. The number of commercial housing projects granted green building certificates can only be counted on the fingers. In Ho Chi Minh City, there are only a few projects such as Ehome 5 of Nam Long, Diamond Lotus Riverside, Rome Diamond lotus of Phuc Khang, The Ascent Thao Dien of Tien Phat, or the North with some projects such as Ecohome, Forest In The Sky, Ecopark… are recognized green properties. The graph below shows that the number of real estate projects complying with LEED (Leadership in Energy and Environmental Design) standards has steadily increased over the years in Vietnam.

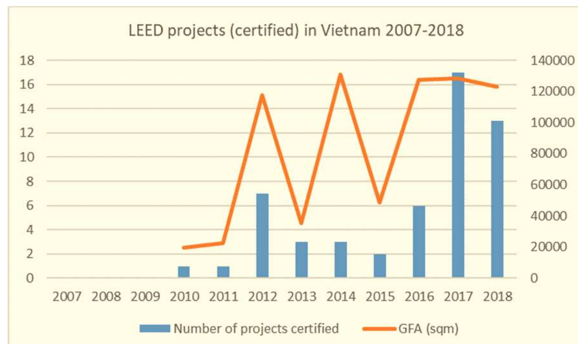


Figure 1: Number of real estate projects complying with LEED

In recent years, urban authorities have recognized the role of urban green space in harmoniously combining natural - human - social factors; improving and enhancing the quality of the living environment and urban landscape, gradually becoming a central goal in urban development planning towards sustainability. Right from the point of view of development, the planning of Hanoi has affirmed that Hanoi is a sustainable "green" city in terms of environment. Like in London (UK), the Hanoi master plan allows only 30% of the land fund to be used for urban development. The remaining 70% of the land fund is for green corridor development. The Green Corridor consists of 40% of conservation areas for high-yield agriculture, biodiversity protection areas and cultural heritage areas. The remaining 30% of the land fund will be used to form development areas based on the conservation of residential areas and craft village activities. This development is controlled and managed to encourage the development of environmentally friendly green activities such as hi-tech agriculture and ecotourism.

4.2. Green Marketing development strategy of real estate enterprises in Vietnam

In order to develop the Green Marketing strategy, especially in the post-COVID-19 period, Vietnam needs to reassess the 4Ps (Product, Price, Promotion, and Place) of the traditional marketing policy to come up with appropriate development and effective.

Green real estate product strategy

Products are the core of meeting the needs of customers. If the product is not good, every attempt by other marketing methods will hardly succeed. Product elements in the Green Marketing strategy are shown through:

Planning, planning to have green real estate requires low construction density, high percentage of trees. Trees are not only one of the elements of nature that play an essential role in the living environment, but also create an aesthetic impression and contribute to the creation of a high-quality living environment, both materially and spiritually, for the inhabitants in urban areas. When the area of green trees and water surface in urban areas reaches 20-50% of the land area, the air temperature can decrease from 3.3 to 3.9 degrees Celsius. The combined effect of shade and evaporation increases when increasing. 25% of the area covered with vegetation can reduce the energy required by 17-57%. Urban trees can reduce 40-50% of solar radiation intensity and absorb 70-80% of solar energy

Design, an environmentally friendly real estate product in the post-COVID-19 era, when the "price escalation" period, all resources and materials are limited to comply with the 3R (Reduction - Reuse - Recycling) principle. (i) Reduction, the product should be designed in such a way that the production can minimize the input materials or be made from materials that minimize the impact on the environment. (ii) Reuse, the real estate materials and equipment can be used multiple times. (iii) Recycling, land, materials, real estate ancillary structures must be recyclable, i.e. capable of being reused, multi-purpose for many different uses, to create new products or have the ability to convert it into raw materials used to create other products

Production, products must be made from clean materials with high productivity, available and easy-to-find materials that do not harm consumers' health as well as adversely affect the environment. Besides, enterprises can also eliminate ineffective production stages, consume a lot of energy, and have negative impacts on the environment; At the same time, businesses can also recycle waste from the production process into input materials.

Eco-labeling, this is a sign that helps consumers to recognize environmentally friendly products. Ecolabel can be created by the real estate Enterprise itself and placed on its products as a commitment to environmental protection. However, usually most businesses assign eco-labeling to their products to independent third parties to ensure honesty and objectivity, and increase the value of customers' trust. This third party may be a non-governmental organization, a government agency, an industry association or another independent company.

Green price strategy

When applying the Green Marketing strategy, real estate businesses need to apply the principle of product pricing based on customer perceptions of product value, not based on product cost. That means that pricing must come from what the customer feels about the product and their willingness to pay for real estate product that provides a benefit of use, health benefits and protective effects environment that is not on the basis of calculating production costs and desired returns.

Accordingly, real estate businesses can price their products through two ways. The first is to set a higher price for green real estate products than for conventional products. At this time, it is required that the company's real estate products need to have superior characteristics, the features that are completely superior to the rest of the product to make customers willing to spend money to buy green products of Enterprise. Generic green product descriptors would not work well with this high pricing approach. The second way is that real estate businesses still sell green real estate products at prices equal to other conventional products. In this case, customers will be willing to spend on eco-friendly products instead of sticking to the one they used to use.

Green promotion strategy

By means of trade promotion, businesses need to create awareness and initial perceptions of customers about green real estate products and services. To ensure the consistency and increase the effectiveness of communication, all conveyed messages need to adhere to the brand positioning, helping customers distinguish the green real estate product of the business from other normal real estate products. The message of green real estate product to customers should be clear so that customers can understand the real estate product's health and environmental protection features, the benefits of green real estate products outperforming other products; avoid hype or exaggeration.

Green distribution strategy

A good distribution policy will have a significant impact on consumers' purchasing decisions, as they will not be able to buy a green real estate product if the real estate product does not appear in a convenient distribution channel or Output is scarce due to enterprises not providing timely (Tien, N.H., et al. 2020).

Green real estate products are often suitable for modern distribution channels such as large openings, sold by reputable agents ... Enterprises can use their distribution channels or cooperate with other partners are committed to protecting the environment. In addition, businesses need to differentiate themselves in the distribution process of their green real estate products compared to conventional products. For example, high quality park planning and construction, committed to using environmentally friendly materials, save fuel and create less emissions...

4.3. Opportunities and challenges for real estate businesses with green marketing

Factors supporting green marketing in Vietnam

Firstly, in general, Vietnam is in the process of industrialization and modernization of the country, the economy has achieved a relatively high growth rate, leading to an improved consumption level of the people. Moreover, the process of economic restructuring and transformation of growth models from breadth to depth, based on factors of productivity and technology that are being prioritized by the Party and the Government are favorable opportunities for Vietnamese real estate businesses to implement green marketing strategies. On the other hand, the State has tightened the regulations on environmental protection, the legal system becomes complete, making real estate businesses under pressure to "green" and also enjoy many incentives when applying policies this book.

Secondly, after COVID-19 has helped increase consumption awareness, the requirements for green real estate products, real estate products that meet safety and quality standards and are environmentally friendly is an essential requirement of the people. This is a potential market for Green Marketing to help businesses exploit.

Along with global warming in recent years, nature's protection mode and the standard of using green real estate products are clearly raised among middle-class and affluent customers in Vietnam. In the real estate market, there have been cases of rapid price increase with green real estate products (typically the Ecopark project in Hung Yen). This strongly stimulates investors to develop towards green real estate. After going through the green marketing process for about 10 years now, a number of real estate businesses have built and matured, perfecting the green marketing capacity for their real estate products.

Third, real estate enterprises' self-awareness about green marketing is increasing. In the post-COVID-19 era, the issue of environmental protection, clean food was widely promoted in domestic media, mentioned in school teaching, commented on at all conferences. From central to local level, building a beautiful image in the eyes of consumers, attracting domestic and international investment capital, at the same time this is a way for businesses to develop sustainable and achieve their goal of increasing long term growth.

Fourth, green marketing strategy is currently a global trend, it is an opportunity for green real estate products to dominate the market of consumers. Simultaneously with the policy of expanding for foreigners to buy houses in Vietnam. In the increasingly fierce competition between real estate businesses and countries in international trade, green products, eco-friendly and eco-friendly products will attract many potential customers around the World. This is also the answer to the very limited real estate export policy in Vietnam.

Fifth, green marketing activities in businesses receive a lot of support from the authorities.

Opportunities and challenges for Vietnamese real estate businesses

First, the biggest challenge for marketers is how to combine the need to "protect the environment" with other basic needs of consumers such as the correlation of cost - efficiency, safety when using, performance, symbolic value and comfort. Vietnamese real estate businesses need to have the right strategy to both meet the needs of consumers and ensure the purpose of environmental protection.

Second, the implementation of the Green Marketing strategy will lead to high investment costs and high real estate product costs. Green marketing is a combination of a series of activities, including modifying products, changing production processes, as well as changing advertising. To implement all stages of green marketing requires real estate businesses to spend a large investment right from the start. Because of the investment in that chain technology, there is an increase in costs and inevitably, the price of the finished product also increases. As a rule, their real estate products will be less competitive in the market. Vietnamese real estate enterprises with small and medium scale, experience as well as scale and cutting related costs such as management costs are currently a headache.

Third, there is no high close coordination between real estate businesses as well as authorities. To change the sense of community in general and consumer consciousness in particular, the essential thing that we need to do is whether we need to know the close association between real estate businesses, the state as well as the strength of the other non-profit organizations. In Vietnam, the legal system on environmental issues is incomplete, standards have not been met by international standards, not practical in Vietnam. Moreover, we have not yet combined the power of non-profit organizations to create a trend that hits strongly on consumer psychology or, if so, that trend has not been active continuously and inevitably has not. There are many results.

Fourth, Vietnamese people like to use cheap goods, do not appreciate the quality of real estate with the green. On the other hand, they are skeptical of advertisements for "green" real estate products because there are many real estate businesses cheating, not transparent ... Therefore, this is a great difficulty for Vietnamese businesses when deploying war green marketing strategy

5. Conclusion and Recommendations

5.1. Conclusion

Through learning and analyzing Green Marketing strategies, real estate businesses need to focus on planning the implementation in the right direction to survive and develop sustainable. The Vietnamese real estate market is developing very fast, with the participation of many economic sectors with increasing competition. Strictly applying Green Marketing strategy with a long-term plan will help real estate businesses pursue profitability, develop sustainable, be socially responsible and embellish their image in the mind of customers, creating an advantage over competitors.

The research has summarized a number of studies related to Green Marketing activities in general to find new points to research in the topic. The author has synthesized the theoretical basis related to the topic including issues: basic green marketing and the current situation of Vietnam after COVID-19; analysis has shown the results achieved and the difficulties that real estate companies face in implementing Green Marketing, analyze the current situation to make effective and appropriate development strategies, point out the achieved results and the difficulties that businesses face in the process of implementing Green Marketing period post COVID-19 in Vietnam.

The main conclusions from the study include:

- Along with the trend of reducing emissions in the world, real estate businesses will be more sustainable if they start green real estate business early, in parallel with implementing green marketing strategies.
- The cost of green real estate and green marketing is large, not suitable for short-term business activities and small businesses.
- To implement green marketing, it is necessary to be based on typical businesses, individuals or products. In which the effectiveness needs to be confirmed over time.

5.2. Recommendation

For real estate businesses, it is necessary to build human resources, especially resources for the Marketing Department. Green marketing is very new in Vietnam so it requires creative and high quality human resources. Real estate Enterprises should focus on three aspects: marketing planning process, budgeting and training people. Promote real estate products, raise environmental awareness, and break consumer behavioral conflicts. Real estate businesses need to properly and honestly promote their products to create trust with customers, helping them to easily choose green real estate products. Create factors that drive consumer buying awareness, including either interest or ethics. Note, there must be a link between green real estate products and consumers 'interests, focusing on environmental factors, but not being able to put environmental factors first and forgetting consumers's benefits.

References

Albino, V., Balice, A. & Dangelico, R.M. (2009) Environmental strategies and green product development: An overview on sustainability–driven companies. *Business Strategy and the Environment*, **18**, 83–96.

Crane, A. (2000) Facing the backlash: Green marketing and strategic reorientation in the 1990s. *Journal of Strategic Marketing*, **8**, 277–296.

Dangelico, R.M., Vocalelli, D. (2017) Green Marketing: An analysis of definitions, strategy steps, and tools through a systematic review of the literature. *Journal of Cleaner Production*, **165**, 1263–1279.

Dean, T.J., Pacheco, D.F. (2014) Green marketing: a strategic balancing act for creating value. *Journal of Business Strategy*, **35**, 14-22.

Davari, A., Strutton, D. (2014) Marketing mix strategies for closing the gap between green consumers' pro–environmental beliefs and behaviors. *Journal of Strategic Marketing*, **22**, 563-586.

Guyader, H., Ottosson, M. & Witell, L. (2017) You can't buy what you can't see: Retailer practices to increase the green premium. *Journal of Retailing and Consumer Services*, **34**, 319-325.

Khoa, H.D., Nhung, T.T.B., (2020) Green logistics application to reduce operating costs for Vietnam logistics enterprises. Journal B&IT, **10**, 2-11.

Ngoc, N.M., Tien, N.H., Anh, D.B.H., (2020) Opportunities and challenges for real estate brokers in post Covid-19 period. *International Journal of Multidisciplinary Research and Growth Evaluation*, **01**, 81-87.

Papadas, K., Avlonitis, G. & Carrigan, M. (2017) Green marketing orientation: Conceptualization, scale development and validation. *Journal of Business Research*, **80**, 236–246.

Peattie, K. (2001) Towards sustainability: The third age of green marketing. *The Marketing Review*, **2**, 129-146.

Peattie, K. (1992) Green marketing M and E Handbooks, Pitman Publishing, England.

Peattie, K., Crane, A. (2005) Green marketing: legend, myth, farce or prophesy? *Qualitative Market Research: An International Journal*, **8**, 357-380.

Polonsky, M.J., Rosenberger, P.J. (2001) Re–evaluating green marketing: a strategic approach. Business Horizons, **44**, 21–30.

Tien, N.H., Ngoc, N.M., Anh, D.B.H., Huong, N.D., Huong, N.T.M., Phuong, T.N.M., (2020) Green marketing development strategy in post Covid-19 period in Vietnam. *International Journal of Multidisciplinary Research and Growth Evaluation*, **01**, 101-106.

Tien, N.H., Ngoc, N.M., (2019) Comparative Analysis of Advantages and Disadvantages of the Modes of Entering the International Market. *International Journal of Advanced Research in Engineering & Management*, **05**, 29-36.

Tien, N.H. (2019) Green economy as an opportunity for Vietnamese business in renewable energy sector. *International Journal of Research in Finance and Management* 2020, **3**, 26-32.

Tien, N.H., Ngoc, N.M., (2019) Related and Non-related Diversification Strategy of Domestic Business Groups in Vietnam. *International Journal of Advanced Research in Engineering and Management,* **5**, 12-17.

Tomasin, L., Pereira, G.M., Borchardt, M. & Sellitto, M.A. (2013) How can the sales of green products in the Brazilian supply chain be increased? *Journal of Cleaner Production*, **47**, 274-282.

Tseng, S., Hung, S. (2013) A framework identifying the gaps between customers' expectations and their perceptions in green products. *Journal of Cleaner Production*, **59**, 174-184.

THE EFFECT OF E-HEALTH-RELATED KNOWLEDGE, PRIVACY & SECURITY CONCERNS ON THE INTENTION TO USE E-HEALTH – A PATIENT'S PERSPECTIVE

Doan, Thi Lien Huong[1] and Tran, Dinh Long[2]

[1] University of Economics, Danang University, 71 Ngu Hanh Son, Danang,

huong.dtl@due.edu.vn

[2] University of Economics, Danang University, 71 Ngu Hanh Son, Danang,

longtd@due.edu.vn

Abstract. E-health was implied as to the application of information & communication technology in the healthcare sector. The importance of e-health was increasingly acknowledged due to its benefits from reducing healthcare costs, providing greater access to healthcare services, especially in the context of COVID-19. Despite the obvious benefit, the intention to use e-health of each individual was prevented/ stimulated by many factors. This study was conducted to identify the influence of e-health-related knowledge, privacy & security concerns on the intention to adopt e-health, with the involvement of 253 patients from the most important medical centers in Danang.

Using the well-known Technology Acceptance Model proposed by Venkatesh & Davis (1996) with modifications regarding e-health related knowledge, privacy & security concerns, the findings indicated that the significant influence of privacy & security concerns on intention to use e-health. Besides, the results also suggested the relationship between e-health-related knowledge and Perceived Usefulness – the predictor of intention to adopt e-health. Based on the research findings, several recommendations were proposed to policymakers and e-health service providers.

Keywords: e-health 1; intention to use 2; TAM 3.

Abbreviation & Description

Abbreviation	Description
PU	Perceived Usefulness
PEOU	Perceived Ease of Use
TAM	Technology Acceptance Model
TRA	Theory of Reasoned Action
P&S	Privacy & Security
ITU	Behavioural Intention to Use

1. Introduction

General Statistics Office of Vietnam (2019) clarified that the ageing index of Vietnam (an indicator that reflecting the increasing proportion of elderly citizen) was roughly tripled since 1999. This situation has happened globally and was referred as population ageing – the swiftly increasing proportion of elderly people in society (World Health Organization, 2018). As a numerous serious health conditions and illnesses associated with this ageing tendency, not only these citizens would not be able to completely experience their later life, but the healthcare system and society as a whole would also have to carry the burden from the increasing cost of providing healthcare services (Carlson et al., 2010).

It has been suggested that the applications of ICT – Information & Communication Technology in healthcare would be significantly helpful in solving above mentioned problems (Steele et al., 2009). These applications was entitled as telemedicine, e-health, or medical informatics, etc. Oh and associates once provided the listing of 51 definition of e-health and related concepts (Oh et al. 2005).

Garshnek et al. (1997) indicated e-health is "the investigation, monitoring and management of patients and the education of patients and staff using systems which allow ready access to expert advice and patient information no matter where the patient or relevant information is located" (p. 38). E- health was also defined as "emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the internet and related technologies" (Eysenbach, 2001).

While there has been few differences among definitions, these two substantial objectives of e-health were shared in common, including (1) providing patients with up-to-date heath information in order to promote the autonomy of patients with their own healthcare (Nicholas et al., 2003); and (2) supporting the interaction of patients and healthcare provider (Baldwin et al., 2002). Mukherjee & McGinnis (2007) once concluded that ultimate purpose of using e-health was to empower patients with latest health information of diagnosis and treatment to support them in making their own decision of healthcare without leaving their houses. Regarding the first goal, it was supported by the 'infomation' aspect of ICT applications such as electronic health record and other related forms of application (Abdekhoda et al., 2018). The second one was fulfilled by the 'communication' aspect with a numerous amount of services as home care/ telecare doctors/ nurses, or online-based consulting service (de Veer et al., 2015).

In the context of healthcare sector, there were two separate end-users of e-health including (1) the professional users (such as general practitioners, nurses, therapists); and (2) the patients. While the study on the e-health utilization of professional users was extensively conducted, the patient's utilization of e-health need further concentration (Holden & Karsh, 2010).

The benefits of e-health and its applications were tremendous, greatly in terms of cost reduction (Sharma et al., 2005), and the greater accessibility to Health-care Services (Burke & Weill, 2005); the decision of individuals to accept, and adopt one specific technology was not that obvious. As a result, the acceptance of ICT application become a prominent topic in healthcare research sector. Recently, the emergence and outbreak of COVID-19 has made the necessitiy of using e-health increase. Therefore, it would be indispensable to elevate a comprehensive understandings on factors that promote or restrain the utilization of e-health.

Consequently, above discussions had contributed the foundation of the following research problem – "what are the factors that effect the patient's intention to use e-health?". As the variety of e-health was tremendous, it would be arduous to investigate the adoption of e-health as a whole. Therefore, the adoption of applications those providing patients with up-to-date heath information as web-based information services, online health record, etc. would be the main focus of this research.

2. Literature Review & Model Development

The acceptance of the technology was acknowledged as the intention to adopt the technology of one individual (Davis, 1993). Due to the spillover of ICT in healthcare recently, the topic of technology acceptance was drawn to the enormous concentration with numerous models that were developed and verified (Dünnebeil et al.2012). Extended from the well-known Theory of Reasoned Action by Aijen & Fishbein (1967), TAM – Technology Acceptance Model was acknowledged as the most popular model that has been applied within this field of research (Holden & Karsh, 2010).

Davis (1989) suggested a modified version of TRA by Fishbein & Ajzen (1975) – TAM (the Technology Acceptance Mode) – for the research field of ICT acceptance. Briggs et al. (2003) described TAM as a model founded on the intention-basis, that explained the eventual use of technology from the end-user perspective.

Figure 1 delivered an overview of the final mode of TAM by Venkatesh & Davis (1996).

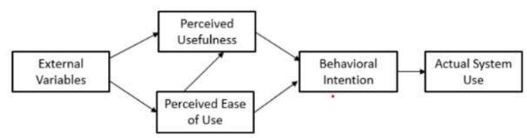


Figure 1. Final Revision of TAM (Venkatesh & Davis, 1996)

Davis (1989) indicated PU – Perceived Usefulness as the degree to which end-users might believe that the use of a (technology) system would benefit their job performance. Since the introduction of TAM, the influence of PU on Behavioral Intention was acknowledged to be significant in numerous previous studies in the sector of healthcare research (Holden & Karsh, 2010). Therefore, in this study, it was supposed that

H1 – PU positively affects Behavioral Intention to Use e-health

PEOU – Perceived Ease of Use was described as "the degree to which a person believes that using a particular system would be free of effort" (Davis, 1989, p. 320). The end-user, or the patient, in specific, with a lack of knowledge in the field of technology, would be resistant to use e-health. By contrast, an e-health application with an intuitive and user-friendly interface would stimulate the intention of the end-user to adopt that e-health application/ system (Abdekhoda et al., 2016). As a result, within this research, it was assumed that

H2 – PEOU positively affects Behavioral Intention to Use e-health

Previous studies suggested that if one system was perceived to be at ease of use, it would enhance the belief of the end-users that using that system might eventually assist their job performance (Holden & Karsh, 2010; Dünnebeil et al., 2012; Abdekhoda et al., 2018). Hence, in this paper, it was expected that

H3 – PEOU positively affects PU of e-health

The construct of Privacy & Security (P&S) was indicated as the extent whether e-health service providers maintained the policies or actions against the breach of personal medical data (Dünnebeil et al. 2012; Abdekhoda et al. 2018). Boddy et al. (2009) indicated that perceived trust in data security positively influenced the PU of e-health.

A recent study also emphasized the concern of privacy & security of medical data as substantial barriers that made end-users frustrated to comprehensively implement e-health (Wang, 2015). The resistance of the end-users would lately not only prevent them from adopting e-health applications (Ford et al., 2016), and but also interfere with the spread of e-health systems (Flaumenhaft & Ben-Assuli, 2018). Therefore, these two following hypotheses were proposed that:

H4 – P&S positively affects the PU of e-health

H5 – P&S positively affects the PEOU of e-health

E-health Knowledge or e-health related knowledge was implicit as the understandings of end-users in using ICT applications for healthcare (Dünnebeil et al. 2012; Rahman et al. 2017). E-health-related knowledge might be referred to as ICT competency (Evans et al., 2018); health literacy (Hemsley et al., 2017); and the capability to exploit the ICT application for fulfilling healthcare purposes (Rahman et al. 2017).

Venkatesh (2000) once indicated that the lack of understandings of a technological system would negatively influence the PEOU of that system. Moreover, users with knowledgeable understandings of a specific system tend to have a greater awareness of the benefit of that system (Kim et al., 2010). Therefore, within this paper, it was assumed that:

H6 – E-health Knowledge positively affects the PU of e-health

H7 – E-health Knowledge positively affects the PEOU of e-health

From the above discussion, the following proposed research model (Figure 2) would be employed in this study.

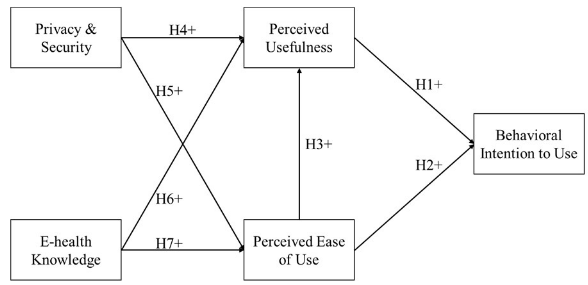


Figure 2. Proposed Research Model

3. Methodology

3.1. Research Instrument

The research model was proposed with six constructs including (1) P&S – Privacy & Security, (2) EhK – E-health knowledge, (3) PU – Perceived Usefulness, (4) PEOU – Perceived Ease of Use, and (5) ITU – Behavioral Intention to Use.

PU, PEOU, and BI were well-defined constructs and were extensively investigated from previous research. The items of these three constructs were thoroughly handpicked from the methodological review by Holden & Karsh (2010) and were made minor adjustments against the context of this study. EhK, P&S were adapted prominently from the study by Dünnebeil et al. (2012), and Abdekhoda et al. (2018). The items of the above-mentioned constructs were measured using a 5-point Likert Scale.

3.2. Data Collection

Target audiences of this study were the end-users or patients in specific. Questionnaires were delivered to the patients involved in treatment from the most important medical centers in Danang. From 500 dispatched questionnaires, 307 questionnaires were sent back, and 253 responses were qualified for data analysis. The characteristics of the sample were indicated in (Figure 3).

Table 1. Sample Characteristics

Criteria	Frequency	Percentage
Gender		
Male	107	43.32%
Female	146	59.11%
Total	**253**	**100.00%**
Age		
15 – 29	63	24.90%
30 – 39	143	56.52%
40 – 54	41	16.21%
>55	6	2.37%
Total	**253**	**100.00%**
Education		
High school	137	54.15%
Undergraduate	81	32.02%
Postgraduate	35	13.83%
Total	**253**	**100.00%**
Income		
< 5 mil VND	10	3.95%
5 – 10 mil VND	134	52.96%
10 – 20 mil VND	79	31.23%
> 20 mil VND	30	11.86%
Total	**253**	**100.00%**

3.3. Data analysis

Data analysis was conducted using Partial Least Square Path Modeling – PLS-PM or Partial Least Square Structural Equation Modeling – PLS-SEM via SmartPLS 3.3.3. This technique, as a variance-based – VB estimator of SEM, was commonly used to explain the recursive, non-recursive, linear, and non-linear structural model (Dijkstra & Schermelleh-Enge, 2014; Dijkstra & Henseler, 2015). PLS-SEM was also a recommended technique of analysis, with prominent accuracy in comparison with regression analysis (Hair et al. 2011). Hair et al. (2014) proposed the process of two-step analysis, in which the measurement model would initially be inspected, and followed by a structural model examination.

4. Research Findings

4.1. Validity of measurement model

The validity of the measurement model should be assessed by both convergent and discriminant validity. While convergent validity implied whether "the indicators belonging to one latent variable measure the same construct", discriminant validity referred to that two different theoretically concepts, represented by two different latent variables, would be statistically different (Benitez et al. 2020).

Table 2. Convergent Validity of Measurement Model

Construct & its indicators	Factor Loading	Cronbach's Alpha	CR	AVE
E-health Knowledge – EhK		**0.899**	**0.937**	**0.832**
EhK1	0.909			
EhK2	0.92			
EhK3	0.907			
Privacy & Security – P&S		**0.864**	**0.908**	**0.711**
P&S1	0.811			
P&S2	0.848			
P&S3	0.873			
P&S4	0.839			
Perceived Ease of Use – PEOU		**0.835**	**0.889**	**0.668**
PEOU1	0.823			
PEOU2	0.847			
PEOU3	0.817			
PEOU4	0.781			
Perceived Usefulness – PU		**0.896**	**0.923**	**0.707**
PU1	0.862			
PU2	0.861			
PU3	0.854			
PU4	0.816			
PU5	0.809			
Behavioral Intention to Use – ITU		**0.829**	**0.897**	**0.745**
ITU1	0.861			
ITU2	0.869			
ITU3	0.859			

Regarding the convergent validity, AVE – average variance extracted & CR – composite reliability are the most popular indicators for examination. Fornell & Larcker (1981) suggested that the value of AVE & CR should be greater than 0.5 & 0.7 correspondingly. The following Figure 4 illustrated the indices of all constructs were significantly acceptable.

In terms of the discriminant validity, three methods commonly used to examine the discriminant validity, including the Fornell-Larcker criterion, the cross-loading criterion, and the HTMT - Heterotrait-Menotrait ratio (Hair et al., 2014). A recent review by Benitez et al. (2020) suggested that the HTMT ratio should be employed to examine the discriminant validity. As can be seen from Figure 3, the overall values of all constructs were significantly lower than 1.0. As a result, the discriminant validity of the measurement model was assured (Voorhees et al., 2016).

Table 3. Discriminant Validity of Measurement Model

	Original Sample	**Sample Mean**	**2.50%**	**97.50%**
ITU -> EhK	0.706	0.707	0.604	0.801
PEOU -> EhK	0.382	0.386	0.248	0.51
PEOU -> ITU	0.534	0.534	0.4	0.657
PU -> EhK	0.607	0.61	0.498	0.707
PU -> ITU	0.718	0.718	0.62	0.816
PU -> PEOU	0.572	0.573	0.45	0.676
P&S -> EhK	0.595	0.595	0.485	0.691
P&S -> ITU	0.63	0.629	0.498	0.746
P&S -> PEOU	0.549	0.549	0.416	0.662
P&S -> PU	0.743	0.743	0.664	0.82

4.2. Structural Paths

Path Coefficients and correspond t-values are synthesized in the following Table 4. t-values were estimated by the Bootstrap Resampling Procedure. As can be seen from Figure 3, 6/7 paths were revealed with corresponding t-values significant at 0.05 level.

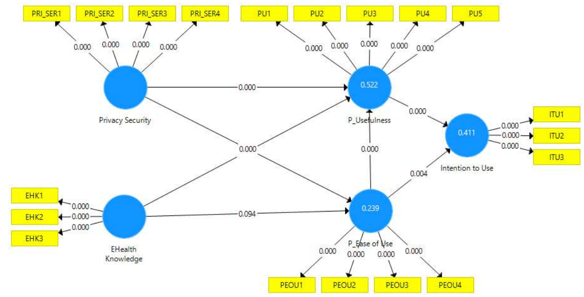


Figure 3. Structural Paths

R2 values indicated that PEOU & PU interpreted 41.1% of the variance in the Behavioral Intention to Use – ITU. While P&S itself explained 23.9% variance of PEOU, P&S, EhK & PEOU jointly interpreted 52.2% of PU.

4.3. Hypothesis testing

Table 4. Bootstrap Results

Hypothesis	**Relationship**	**ß**	**t-value**	**p-value**	**VIF**	**Comment**
H1	PU->ITU	0.526	8.993	0	1.34	Supported
H2	PEOU->ITU	0.188	2.917	0.004	1.34	Supported
H3	PEOU->PU	0.218	3.937	0	1.313	Supported
H4	P&S->PU	0.416	7.183	0	1.605	Supported
H5	P&S->PEOU	0.41	5.601	0	1.385	Supported
H6	EhK->PU	0.255	4.255	0	1.406	Supported
H7	EhK->PEOU	0.127	1.677	0.094	1.385	Rejected

Perceived Usefulness was revealed as the factor that conserved the prominent impact on Behavioral Intention to Use (ß = 0.526, t = 8.993). Data analysis indicated that Privacy & Security was posited significant influence on Perceived Usefulness and Perceived Ease of Use with (ß = 0.416, t = 7.183) & (ß = 0.41, t = 5.601) accordingly. Regarding the E-health knowledge, the result indicated its effect on Perceived Usefulness with (ß = 0.255, t = 4.255). The effect of E-health Knowledge on Perceived Ease of Use was not supported by the research findings.

5. Discussions

The research findings foremost confirmed the relationship of Perceived Usefulness and Behavioral Intention to Use (H1 – supported). This relationship was widely approved and enlisted in extensively previous research (Lai, 2017; Holden & Karsh, 2010). The result has also affirmed the influence of Perceived Ease of Use on Perceived Usefulness and Behavioral Intention to Use (H2, H3 – supported). These findings were shared in common with the previous studies in the field of healthcare (Holden & Karsh, 2010; Dünnebeil et al., 2012; Abdekhoda et al., 2018).

Regarding the factor of Privacy & Security, this study uncovered that Privacy & Security was a crucial predictor of both Perceived Usefulness & Perceived Ease of Use (H4, H5 – supported). The findings were conforming to the results of the previous study including Dünnebeil et al. (2012); Abdekhoda et al. (2018); and Khan et al. (2019). In terms of e-health knowledge, this construct referred to the understandings of end-users in using ICT applications for healthcare. This component was examined to conserve the considerable influence on both Perceived Usefulness & Perceived Ease of Use (Dünnebeil et al., 2012; Rahman et al., 2017; Ghaddar et al., 2020).

In terms of e-health knowledge, the findings revealed the significant influence of EhK on PU (H6 – supported). This result was also supported by previous investigations (Cebeci et al., 2019; Kim et al., 2010). However,within this study the relationship between EhK and PEOU was not statistically supported. The findings de facto contradicted the previous research that emphasized the significant relationship of EhK & PEOU (Cebeci et al., 2019; Dünnebeil et al., 2012). This outcome might be explained by the fact that the e-health systemư/ applicationư in Vietnam were in the developing stage and the complexity of e-health was not extensive (KPMG, 2020). Another suggestion to explain this contradiction would be the user interface of e-health system/ application. While the interface of system & application employed by doctor or practioner was complicated (Dünnebeil et al., 2012), the one utilized by patient would be easier to access. As a result, the end-user (patients in specific) might not be aware of this relationship

6. Conclusions

This study was conducted to answer the research question that "what are the factors that affect the patient's intention to use e-health?". Generally speaking, the overall purpose of this study was fulfilled. This study was once again affirmed the influence of Perceived Usefulness and Perceived Ease of Use on Behavioral Intention to Use. Moreover, the research findings also supported the effect of Perceived Ease of Use on Perceived Usefulness. Besides, this research has uncovered the role of Privacy & Security and E-health Knowledge on Behavioral Intention to Use via PU and PEOU.

Regarding Privacy & Security, research findings illustrated a significant effect on PU and PEOU, which lately increase the degree of e-health acceptance. The findings suggested that in addition to building a system/ application with a user-friendly interface, e-health service providers (both public & private) should focus on designing a highly secured e-health system/ application against the virus, malware and other threat. Additionally, policy regarding the data security and the privacy of personal information should be advisedly constructed and stayed in line with existent regulation of government. Furthermore, the government should actively involve in regulating the extent regarding the Privacy & Security of e-health.

In terms of E-health Knowledge – the capability of end-user to exploit ICT for healthcare purposes, the study revealed that this construct significantly impacted the perceived usefulness – the most important predictor of Behavioral Intention to Use. From the findings, it would be highly recommended that e-health providers should pay their attention to raising the awareness of citizen on the existence of e-health, along with its benefits. The solutions might come from establishing an attractive launching campaign of new e-health system/ applications; or providing a visual manual guide for utilizing e-health system & application, etc.

In conducting this study, some limitations needed to be mentioned. Initially, the number of respondents was still humble and it would be recommended that later research would be conducted with a greater number of samples. Then, as the PU, and PEOU explained 41% of the variance of behavioral intention to use, there were other influencing factors on ITU that needed to be addressed in later studies. Lastly, it was astonishing that the relationship between e-health knowledge and PEOU was not supported within this study. As this relationship was extensively posited in the previous study (Dünnebeil et al., 2012, Rahman et al., 2017), the upcoming study should focus on re-examining this relationship.

7. Acknowledgment

This research is funded by Funds for Science and Technology Development of The University of Danang under project number B2018-ĐN04-12.

References

Abdekhoda, M., Dehnad, A., & Khezri, H. (2018). The effect of confidentiality and privacy concerns on adoption of personal health record from patient's perspective, *Health and Technology*, 9, 463–469.

Abdekhoda, M., Dehnad, A., Mirsaeed, S.J.G., & Gavgani, V.Z. (2016). Factors influencing the adoption of E-learning in Tabriz University of Medical Sciences, *Medical Journal of the Islamic Republic of Iran*, 30 (457), 1-7.

Baldwin, L., Clarke, M., & Jones, R. (2002). Clinical ICT systems: augmenting case management, *Journal of Management in Medicine*, 16(2/3), 188-198.

Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research, *Information & Management*, 57, 1-20.

Briggs, R.O., Vreede, G.J.D., & Nunamaker, J.F. (2003). Collaboration engineering with ThinkLets to pursue sustained success with group support systems, *Journal of Management Information Systems*, 19 (4), 31-64.

Boddy, D., King, G., Clark, J., Heaney, D., & Mair, F. (2009). The influence of context and process when implementing e-health, *BMC Medical Informatics and Decision Making*, 9 (1), 9, https://doi.org/10.1186/1472-6947-9-9.

Burke, L., & Weill, B. (2005). *Information Technology for the Health Professions,* 2nd ed, PEARSON Prenctice Hall, Uppder Saddle River: NJ.

Carlson, J.J., Sullivan, S.D., Garrison, L.P., Neumann, P.J., & Veenstra, D.L. (2010). Linking payment to health outcomes: A taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers, *Health Policy*, 96(3), 179-190.

Cebeci, U., Ince, O., & Turkcan, H. (2019). Understanding the Intention to Use Netflix: An Extended Technology Acceptance Model Approach, *International Review of Management and Marketing*, 9(6), 152-157.

Davis, F.D. (1989), Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, 13(3), 319-340.

Davis, F.D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts, *International Journal of ManMachine Studies*, 38(3), 475-487.

de Veer, A.J.E., Peeters, J.M., Brabers, A.E.M., Schellevis, F.G., Rademakers, J.J.D.J.M., & Francke, A.L. (2015). Determinants of the intention to use e-Health by community dwelling older people, *BMC Health Services Research*, 15:103.

Dijkstra, T., & Henseler, J. (2015). Consistent partial least squares path modeling, *MIS Quarterly*, 39 (2), 297–316.

Dijkstra, T., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models, *Psychometrika*, 79(4), 585–604.

Dünnebeil, S., Sunyaev, A., Blohm, I., Leimeister, J.M., & Krcmar, H. (2012). Determinants of physicians' technology acceptance for e-health in ambulatory care, *International Journal of Medical Informatics*, 81, 746-760.

Evans, B.A., Beverly, C.J., Tsai, P-F., Rettiganti, M., Lefler, L.L., & Parks, R.F. (2018). Older Adults' Live Demonstration of Electronic Personal Health Record Use: Factors Mediating Initial Proficiency, *Computers, Informatics, Nursing*, 36(12), 603-609.

Eysenbach, G. (2001). What is e-health?, *Journal of Medical Internet Research*, 3(2):e20, doi: 10.2196/jmir.3.2.e20.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*, Reading, MA: Addison-Wesley.

Flaumenhaft, Y., & Ben-Assuli, O. (2018). Personal health records, global policy and regulation review, *Health Policy*, 122(8), 815-826.

Ford, E.W., Hesse, B.F., & Huerta, T.R. (2016). Personal Health Record Use in the United States: Forecasting Future Adoption Levels, *Journal of Medical Internet Research,* 18(3):e73, doi:10.2196/jmir.4973.

Fornell, C.G., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research*, 18(1), 39–50.

Garshnek, V., Logan, J.S., & Hassell, L.H. (1997). The telemedicine frontier: going the extra mile, *Space Policy*, 13(1), 37-46.

General Statistics Office (2019). *Completed Results of the 2019 Viet Nam Population and Housing Census*, Statistical Publishing House, Hanoi – Vietnam.

Ghaddar, S., Vatcheva, K.P., Alvarado, S.G., & Mykyta, L. (2020). Understanding the Intention to Use Telehealth Services in Underserved Hispanic Border Communities: Cross-Sectional Study, *Journal of Medical Internet Research*, 22(9), e21012, 1-13, doi: 10.2196/21012.

Hair Jr, J.F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V.G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research, *European Business Review*, 26(2), 106-121.

Hair, J.F., Ringle, C.M. & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet, *The Journal of Marketing Theory and Practice*, 19(2), 139–152, doi: 10.2753/MTP1069-6679190202.

Hemsley, B., Rollo, M., Georgiou, A., Balandin, S., & Hill, S. (2018). The health literacy demands of electronic personal health records (e-PHRs): An integrative review to inform future inclusive research, *Patient Education and Counseling*, 101(1), 2-15.

Holden, R.J., & Karsh, B.T. (2010). The technology acceptance model: its past and its future in health care, *Journal of Biomedical Informatics*, 43 (1), 159–172.

Khan, I., Xitong, G., Ahmad, Z., & Shahzad, F. (2019). Investigating Factors Impelling the Adoption of e-Health: A Perspective of African Expats in China, *Healthcare Informatics*, 1-12, https://doi.org/10.1177/2158244019865.

Kim, C., Mirusmonov, M., & Lee, I. (2010). An empirical examination of factors influencing the intention to use mobile payment, *Computers in Human Behavior*, 26(3), 310-322.

KPMG & OUCRU. (2020). Digital Health in Vietnam – Marketing Intelligence Report, Retrieved from https://assets.kpmg/content/dam/kpmg/vn/pdf/publication/2021/digital-health-vietnam-2020-twopage.pdf. (*accessed on June 13th, 2021*).

Lai, P.C. (2017). The literature review of technology adoption models and theories for the novelty technology, *Journal of Information Systems and Technology Management*, 14(1), 21-38.

Mukherjee, A., & McGinnis, J. (2007). E-healthcare: an analysis of key themes in research, *International Journal of Pharmaceutical and Healthcare Marketing*, 1(4), 349-363.

Nicholas, D., Huntington, P., Williams, P., & Gunter, B. (2003). First steps towards providing the UK with health care information and advice via their television sets: and evaluation of four Department of Health sponsored pilot services, *Aslib Proceedings*, 55(3), 138-154.

Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What Is eHealth: A Systematic Review of Published Definitions, *Journal of Medical Internet Research,* 7(1):e1, doi:10.2196/jmir.7.1.e1.

Rahman, M.S., Hanifi, S., Khatun, F., et al. (2017). Knowledge, attitudes and intention regarding mHealth in generation Y: evidence from a population based cross sectional study in Chakaria, Bangladesh, *BMJ Open*, 7:e016217, doi:10.1136/bmjopen-2017-016217.

Sharma, S.K., Ahmed, N., & Rathinasamy, R.S. (2005). E-healthcare: a model on the offshore healthcare delivery for cost saving, *International Journal of Healthcare Technology and Management*, 6(3), 331-351.

Steele, R., Lo, A., Secombe, C., & Wong, Y.K. (2009). Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare, *International Journal of Medical Informatics*, 78(12), 788-801.

Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model, *Information Systems Research*, 11(4), 342-365.

Venkatesh, V., & Davis, F.D. (1996). Model of Antecedents of Perceived Ease of Use, *Decision Sciences*, 27(3), 451-481.

Voorhees, C.M., Brady, M.K., Calantone, R., Ramirez, E. (2016). Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies, *Journal of the Academy of Marketing Science*, 44(1), 119–134.

Wang, C-K.(2015). Security and privacy of personal health record, electronic medical record and health information, *Problems and Perspectives in Management*, 13(4), 19-26.

World Health Organization (2018). Ageing and health. Retrieved from https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. (*accessed on June 13th, 2021*).

APPLICATION OF CAPACITATED VEHICLE ROUTING PROBLEM WITH MULTI-DEPOT AND MULTI-TRIP: A CASE STUDY OF WINE RETAILER IN VIETNAM

Le Tran Phuong Anh[1], *Nguyen Hoang Huy[2] , *Do Vinh Truc[3]

School of Industrial Engineering and Management

The International University - Vietnam National University, Vietnam

Email: [1]ltpa.209@gmail.com, [2]nhhuy@hcmiu.edu.vn, [3]dvtruc@hcmiu.edu.vn

Abstract. In recent years, the Logistics Industry in Vietnam has earned significant revenue for the country's economy and led to the expanding in other disciplines. However, undeveloped infrastructure and outdated administration systems have restricted the sector's competitiveness, led to high cost and a considerable bottleneck. Different factors issues distinguished cases, but a practical delivery distribution of a wine retailer in the region will be considered in this paper, the one that addresses the problems related to capacitated vehicle routing problems (CVRP) with multi-depot, multi-trip, and a heterogeneous fleet of transportations for a single commodity. The study aims to find optimal delivery routes with optimal travelling costs regarding the constraints of vehicle availability and cargo loads, combined with capacity issues arising from vehicles' originated distribution centers. Furthermore, the paper also expands to the upper echelon of the distribution network by developing an extended model to intervene in the company's inbound shipments. The models are referred using mixed integer programming and treated by the advance of commercial optimization packages, which promises to generate exact solutions in a reasonable computation time. The computational results also present that the proposed formulation is competitive against the original instances from literature, in which the constraints expand to almost all practical perspectives that the case is currently facing.

Keywords: Capacitated Vehicle Routing with Multi-trip, Multi-depot, Heterogeneous Fleet of Vehicle, Partial Shipment.

1. Introduction

1.1. Wine Retailer in Vietnam

In recent years, the buying culture, in both domestic and worldwide scale, has been reported for specific and notable changes, such as the refinement in pricing categories, wider and more accessible product distribution, and the growth of middle-end and high-end products market. As more wealthy people are, the need of upgrading living standards and showing their social status are more covetous, contributing to the increment of upmarket goods in customer's shopping patterns. Unlike necessity goods, which can be purchased repetitively to serve daily basic needs, upmarket goods or superior goods are produced in limited quantities and mostly in the brand's original place of birth. Thereby, such markets not only relied on the improvement of transport infrastructure, but also the foreign commercial policies to ease the process of import – export custom. Those practices, fortunately, have been implemented or under-negotiated by the Vietnam government as an effort to satisfy their citizen's buying demands for luxury items.

According to Nielsen, a survey shows that Vietnam positions third within the world in terms of affection for branded merchandise, as it was outperformed by China and India. This left the opportunities for several luxury brands to set their footprints into Vietnam market, which can be named as Jaguar, Bentley, Lamborghini for car industry; Channel, Gucci, Louis Vuitton, Prada for apparel industry; Cartier, Rolex, Hermès, Tiffany & Co. for accessories industry, etc. However, it is worth mentioning that those products are mostly imported, and are distributed by presentative retailers such as DAFC, Maison, IPPG, etc. and in some cases, the brand will establish their own pop-up or mono brand stores in Vietnam regardless of any intermediators. Their vanguard investments have reaped achievements which include the thriving in consumption and profits, the expanding in stores and branches, but the most critical one is the push for other luxury industries to dive in the market. Besides the three fields mentioned above, over the past few years, there are two more disciplines that have emerged in the Vietnam economy, which are luxuries service (e.g., pampering treatments, red carpet events, high-class performances, etc.) and wine industry. Although the terms are quite new, those yields have marked a significant reputation and risen in sales. Nevertheless, both management and distribution systems in the two segments are assessed as amateur and insufficient, leaving the inspiration and motivation for the researchers to conduct the topic.

This paper reports the results of applying Capacitated Vehicle Routing Problems model (CVRP) to solve the case of three-echelon distribution of wine retailer named Retailer R Company – one of the biggest retailers in the field up till now. Detailed problems of the case study will be introduced in the next section, followed by the desired achievements, and any assumptions if needed. Secondly, an overview of the papers related to CVRP topic and others in VRP field will be synthesized. Thirdly, a mathematical programming formulation will be presented for the problem, as well as its solving procedure. Fourthly, a numerical result and its visualization will be shown. Conclusions and recommendations will be listed out in the final part. Lastly, due to the privacy policy of the company, "Retailer R" will be addressed instead of using the real business name.

1.2. Problem statement

Apparently, it is a need for any wine entrepreneurs in Vietnam to optimize their distribution system and its related factors. In our case, the objectives will focus on Retailer R's root problems as follow:

The distribution network system of the company currently includes three elements: (1) Two sources of clients scattering mostly in Ha Noi and Ho Chi Minh City. (2) One Distribution Center (DC) located in Ha Noi and Two DCs in Ho Chi Minh City. (3) One warehouse located in Da Nang City.

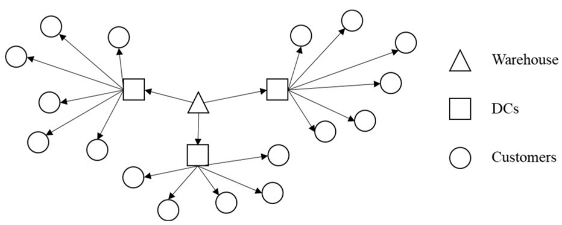


Figure 1. Distribution Network

The transportation sources are various in types and capacity. To be specific, there are four fleets of vehicles in the company: (1) The first one belongs to the warehouse, including large capacity trucks; (2) The three remaining fleets belong to each distribution center, including trucks and mini trucks. From wine factories to warehouse, vessels are used; however, the paper will only consider national distribution network.

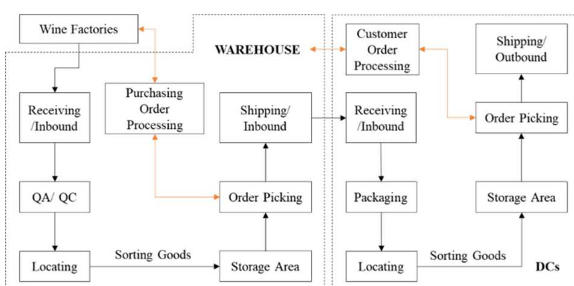


Figure 2. Current Process in Distribution Network System

Accordingly, the current facilities and equipment cause real-case conflicts and drawbacks in operation, composing:

For routing factor, vehicle utilization and assignment of routes are the main concern since the shipments are made based on the subjective acquisitions of operators. Thereby, the delivery patterns are extremely intricacy: some are direct shipments that lead to significant raise in travelling and respective back-haul costs; some are integrated without any standards that lead to a long driving distance which cause fatigue for drivers and put vehicles in overused status. Moreover, for the shipment that has demand smaller than vehicle capacity, it will create a waste in utilization and low productivity.

For inventory factor, insufficient operations also rise at DCs since replenishments from warehouse are made only when DCs stock out (in figure 2, after picking orders for outbound clients, if the operator notices that there is no stock left, then replenishment orders will be sent to warehouse). This practice first causes a bottleneck in order processing, as DC's operators must wait for the restocks from the warehouse. Second, since the current DC capacity cannot cover the demand of all regional clients at once, bottleneck in packaging and storage of goods will be an issue in the case that outbound products surpass inbound ones.

1.3. Objectives

Addressing those problems, this paper aims to achieve the foremost objective: to minimize all costs in view of transportation for the whole process covering from warehouse to clients, and simultaneously disentangle the constraints related to inventory and routing issues with following goals:

(1) To satisfy the demand of all customers making orders in the season.

(2) To ensure the inventory level does not exceed the capacity (inventory decision).

(3) To utilize transportation in an effective way: integrate orders to optimize vehicle's travelling distance if possible and must consider capacity issues; assign multi-trip approach to overcome the limitation of resources; apply partial shipment policy if needed (routing decision).

1.4. Scope and limitations

One of the characteristics of luxury items is they are usually purchased in small quantities but with high value. Thereby, in Retailer R case, some demands are relatively small. Moreover, the paper aims to apply models only on the distributing and selling phase in the business process, from warehouse to distribution centers and clients, throughout the whole country. The purchasing phase where Retailer R makes orders from foreign wine farms/ businesses will be excluded. However, with the aim of expressing the challenges comprehensively, data in the research is chosen from different periods in the year including low demand season, normal demand season and special season such as Tet holiday. Lastly, since the products distributed by Retailer R are not much different in size and weight, all categories will be assumed as one type with a unit is a bottle.

2. Literature review

Capacitated Vehicle Routing Problem is a real-life obstacle in which multiple clients are assigned contemporaneously to individual transportation to keep the total traveling distance and costs as least as conceivable. Moreover, this practice also improves in capacity utilization of vehicles, where the cargo load for each trip will be optimized uniformly with the moving routes.

Various research is investigated to solve VRP in the last few decades, most of them are expanded and combined with add-in perspectives that attached interdependently with VRP, which are location and inventory. Thereby, this field is seen to be applied in many industries, from manufacturing to trading businesses. Deng et al. (2016) studied a closed-loop supply chain in the e-commerce industry. They proposed a heuristic inspired from Hybrid Ant Colony Optimization Algorithm to solve a nonlinear programming model considered defects returns. Rahbari et al. (2020) presented a linear mixed-integer programming model for a five-echelon red meat supply chain in Iran and conducted results with the effort of a general algebraic modeling system software. Based on the incorporation of Adaptive Genetic Algorithm and Simulated Annealing (SA), Guo et al. (2018) developed and solved a NIP model of a closed-loop supply chain case. Considering stochastic demand in the e-commerce industry, Liu et al. (2015) integrated a Pseudo-Parallel Genetic Algorithm and SA to solve an NLP model.

The distribution network problem also extends to many perspectives such as considering back haul trips, time windows for inventory policies or targeting to service levels, etc. Amongst those, multi-trip vehicle routing (MT-VRP) has attracted the attention of several researchers. For instance, the first study belonged to Salhi and Sari (1997) who proposed a multi-level heuristic to simultaneously deal with allocation, delivery trips and vehicle fleets problems. Addressing overtime issues in multi-trip, Petch and Salhi (2003) attempted to develop a multi-phase constructive heuristic algorithm. Plus, Wassan et al. (2017) aimed to enlarge research avenues by combining back haul trips to the original MT-VRP model and solving by Two-Level VNS methodology.

However, hardly any papers in the literature considered luxury products. To the extent of knowledge, this paper is one of those rare articles that research wine products distribution problems. An CVRP model reviewing inventory capacity (Rahbari et al. 2020); multi-trip, multi-depot obstacle (Wassan et al. 2017)(Ambrosino and Grazia Scutellà 2005)(Zhen et al. 2020); a heterogeneous fleet of vehicles; and partial shipment, which inspired from previous articles will be proposed and adjusted to disentangle the current challenges for the Retailer R Company case.

3. Methodology

3.1. Mathematical model

In this section, the introduced problems will be stated in detail and translated into mathematical form. Before formulating the model, some assumptions are made as follow:

- All products are assumed to be homogeneous (unit: one bottle) and always available at the warehouse.
- Each trip begins and ends at the same distribution centers/ warehouse.
- The total demand of customers served on one trip cannot surpass the vehicle's capacity.
- The inbound and outbound goods must be balanced and cannot exceed the distribution center's capacity.
- The capacity of the warehouse is finite.
- No service times (loading/ unloading) are considered.
- Time unit is monthly.
- Vehicle unit is one truck.

There will be two models in this paper. Both objectives are the same, which is to optimize the transportation cost. However, the constraints are much different since the first model will be implemented for DC to customers phase while the other will be used for warehouse to DC phase.

3.1.1. Notation

Indices

i, j, l *Index of customer*
d *Index of distribution center (DCs)*
k, v *Index of vehicle*
w, t *Index of trip*
p *Index of period*

Sets

Model 1

D *Set of distribution centers*
K *Set of transportations available at each DC*
N *Set of customers*
P *Set of periods*
W *Set of trips for each vehicle at DCs*

Model 2

T *Set of trips for each vehicle at warehouse*
V *Set of transportations available at warehouse*

3.1.2. Parameters

Model 1

$cap^1_{d,k}$ *Capacity of vehicles at each DC*
cap^2_d *Storage capacity of each DC*
$cos^1_{d,k}$ *Transportation cost per one kilometer of vehicle k at each DC*
$dem^1_{p,i}$ *Demand of customer i in period p*
$dis^1_{i,j}$ *Distance between customer with each other*
$dis^2_{d,i}$ *Distance between each DC d to customer i*

m Maximum distance that vehicle can operate in one trip.

Model 2

cap_v^2 *Capacity of vehicles at warehouse*
cos_v^2 *Transportation cost per one product of vehicle v at warehouse*
$dem_{p,d,t}^2$ *Demand of DCs i for trip t in period p (replenishment orders)*
dis_d *Distance between warehouse and DCs*
$qs_{p,d,t}$ *Quantity of goods shipped to DCs i for trip t in period p.*

3.1.3. Decision variables

Model 1

$R_{p,d,k,w}$ *Cargo load of vehicle k when operating trip w in period p*

$X_{p,i,d,k,w}$ *Binary variable,* $X_{p,i,d,k,w} = \begin{cases} 1 \text{ if customer } i \text{ is assigned to trip } w \text{ of vehicle } k \text{ at DC } d \text{ in period } p \\ 0, \text{ otherwise} \end{cases}$

$Y_{p,i,j,d,k,w}$ *Binary variable,* $Y_{p,i,j,d,k,w} = \begin{cases} 1 \text{ if customer } i \text{ precedes customer } j \text{ in trip } w \text{ of vehicle } k \text{ at DC } d \\ 0, \text{ otherwise} \end{cases}$

$Z_{p,d,k,w}$ *Binary variable,* $Z_{p,d,k,w} = \begin{cases} 1 \text{ if vehicle } k \text{ is assigned to trip } w \text{ in period } p \\ 0, \text{ otherwise} \end{cases}$

$FC_{p,i,d,k,w}$ *Binary variable,* $FC_{p,i,d,k,w} = \begin{cases} 1 \text{ if customer } i \text{ is the first customer in trip } w \text{ of vehicle } k \text{ at DC } d \\ 0, \text{ otherwise} \end{cases}$

$LC_{p,i,d,k,w}$ *Binary variable,* $LC_{p,i,d,k,w} = \begin{cases} 1 \text{ if customer } i \text{ is the last customer in trip } w \text{ of vehicle } k \text{ at DC } d \\ 0, \text{ otherwise} \end{cases}$

Model 2

$S_{p,d,v,t}$ *Binary variable,* $S_{p,d,v,t} = \begin{cases} 1 \text{ if DC } d \text{ is assigned to trip } t \text{ of vehicle } v \text{ in period } p \\ 0, \text{ otherwise} \end{cases}$

$L_{p,d,v,t}$ *Load of vehicle v traveling to DC d for trip t in period p*

3.2. Model

3.2.1. Objective function

Model 1

$$Minimize\ R1 = \sum_{p\in P}\sum_{i,j\in N}\sum_{d\in D}\sum_{k\in K}\sum_{w\in W} dis_{i,j}^1 * cos_{d,k}^1 * Y_{p,i,j,d,k,w} + \sum_{p\in P}\sum_{i\in N}\sum_{d\in D}\sum_{k\in K}\sum_{w\in W} dis_{d,i}^2 * cos_{d,k}^1 * FC_{p,i,d,k,w} + \sum_{p\in P}\sum_{i\in N}\sum_{d\in D}\sum_{k\in K}\sum_{w\in W} dis_{d,i}^2 * cos_{d,k}^1 * LC_{p,i,d,k,w}$$

Model 2

$$Minimize\ R2 = \sum_{p\in P}\sum_{d\in D}\sum_{v\in v}\sum_{t\in T} S_{p,d,v,t} * dis_d * cos_v^2$$

3.2.2. Constraints

Model 1

$$\sum_{d\in D}\sum_{k\in K}\sum_{w\in W} X_{p,i,d,k,w} = 1 \quad \forall\, p\in P, \forall\, i\in N \qquad (1)$$

$$FC_{p,l,d,k,w} + \sum_{i\in N} Y_{p,i,l,d,k,w} = X_{p,l,d,k,w} \quad \forall\, p\in P, \forall\, l\in N, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (2)$$

$$LC_{p,l,d,k,w} + \sum_{j\in N} Y_{p,l,j,d,k,w} = X_{p,l,d,k,w} \quad \forall\, p\in P, \forall\, l\in N, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (3)$$

$$FC_{p,l,d,k,w} \le \sum_{i\in N} Y_{p,l,i,d,k,w} \quad \forall\, p\in P, \forall\, l\in N, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (4)$$

$$LC_{p,l,d,k,w} \le \sum_{j\in N} Y_{p,j,l,d,k,w} \quad \forall\, p\in P, \forall\, l\in N, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (5)$$

$$Y_{p,i,j,d,k,w} = 0 \quad \forall\, p\in P, \forall\, i,j\in N: i=j, \forall\, d\in D, \forall\, k\in K \ \forall\, w\in W \qquad (6)$$

$$Z_{p,d,k,w} = \sum_{i\in N} FC_{p,i,d,k,w} \quad \forall\, p\in P, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (7)$$

$$Z_{p,d,k,w} = \sum_{i\in N} LC_{p,i,d,k,w} \quad \forall\, p\in P, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (8)$$

$$Z_{p,d,k,w+1} \le Z_{p,d,k,w} \quad \forall\, p\in P, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (9)$$

$$Z_{p,d,k+1,w} \le Z_{p,d,k,w} \quad \forall\, p\in P, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (10)$$

$$R_{p,d,k,w} = \sum_{i\in N} X_{p,i,d,k,w} * dem_{p,i}^1 \quad \forall\, p\in P, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (11)$$

$$R_{p,d,k,w} \le Z_{p,d,k,w} * cap_{d,k}^1 \quad \forall\, p\in P, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (12)$$

$$\sum_{k\in K} R_{p,d,k,w} \le cap_d^2 \quad \forall\, p\in P, \forall\, d\in D, \forall\, w\in W \qquad (13)$$

$$\sum_{i\in N} dis_{d,i}^2 * FC_{p,i,d,k,w} + \sum_{i\in N} dis_{d,i}^2 * LC_{p,i,d,k,w} + \sum_{i\in N}\sum_{j\in N} dis_{i,j}^1 * Y_{p,i,j,d,k,w} \le m \quad \forall\, p\in P, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (14)$$

$$Z_{p,d,k,w} \in \{0,1\};\ Y_{p,i,j,d,k,w} \in \{0,1\};\ X_{p,i,d,k,w} \in \{0,1\};\ FC_{p,i,d,k,w} \in \{0,1\};\ LC_{p,i,d,k,w} \in \{0,1\} \quad \forall\, p\in P, \forall\, i,j\in N, \forall\, d\in D, \forall\, k\in K, \forall\, w\in W \qquad (15)$$

The objective function is to minimize the travelling cost of all shipments arising from distribution centers to customers. Constraint (1) implies that each customer is served exactly by one distribution center, one vehicle and one trip in one specific period. Constraint (2) and (3) restrict the condition of the first and last customer in one trip. For example, if l is assigned to trip w of vehicle k at DC d and l is the first customer ($X_{p,l,d,k,w} = 1$ and $FC_{p,l,d,k,w} = 1$), the other variable in the equation ($Y_{p,i,l,d,k,w}$) must equal 0, which means there will be none of any customers i precedes customer l, vice versa for the last customer constraint. Constraint (4) – (6) remove the redundant arcs: travel from DC to customer i only and comeback; travel from customer i to customer j and return to customer i. Constraint (7) and (8) state that if trip w by vehicle k is performed, there must be only one first and last customer. Constraint (9) bounds that trip $w + 1$ of vehicle k can only be created if trip w has been operated. Constraint (10) warrants that all vehicles should be utilized. Constraint (11) ensures that the initial cargo load is equal the sum of demand of all customers in one specific trip w operated by vehicle k at DC d. Constraint (12) addresses that the cargo load is smaller than vehicle k capacity if it is chosen to operate trip w. Constraint (13) means that the total cargo load of all vehicles k traveling in trip w (outbound goods) can not exceed the capacity of DC where those vehicles belong to. This variable can also be understood as replenishment quantity. Constraint (14) declares that the total distance that vehicle k traveling in trip w is smaller than the maximum distance that the government allows drivers to drive in a day. Constraints (15) define the format of decision variables.

__________________*Model 2*__________________

$$qs_{p,d,t} = dem^2_{p,d,t}$$
$\forall\, p \in P, \forall\, d \in D, \forall\, t \in T$ *(16)*

$$\sum_{v \in V} S_{p,d,v,t} \geq 1$$
$\forall\, p \in P, \forall\, d \in D, \forall\, t \in T$ *(17)*

$$\sum_{d \in D} S_{p,d,v,t} \leq 1$$
$\forall\, p \in P, \forall\, d \in D, \forall\, t \in T$ *(18)*

$$qs_{p,d,t} = \sum_{v \in V} L_{p,d,v,t}$$
$\forall\, p \in P, \forall\, d \in D, \forall\, t \in T$ *(19)*

$$L_{p,d,v,t} \leq S_{p,d,v,t} * cap^2_v$$
$\forall\, p \in P, \forall\, d \in D, \forall\, v \in V, \forall\, t \in T$ *(20)*

$$L_{p,d,v,t} \geq 0;\ S_{p,d,v,t} \in \{0,1\}$$
$\forall\, p \in P, \forall\, d \in D, \forall\, v \in V, \forall\, t \in T$ *(21)*

The objective function is to minimize the travelling cost of all shipments arising from warehouse to distribution centers. Constraint (16) ensures the flow of products moving out and in DCs are equal. Constraint (17) means that if the demand of DC for one trip surpasses the capacity of one vehicle, the system will separate the replenishment order into multiple shipments operated by different transportations. Constraint (18) prevents overlap of resources by restricting each vehicle to be used at most one time for one replenishment trip. Constraint (19) warrants that the total cargo load of every vehicle visiting DC d in trip t is equal to the total required demand of that DC (inbound goods). Constraint (20) restricts that the cargo load of trip t by vehicle v is not exceeded its capacity. Constraint (21) defines that format of decision variables.

3.3. Data collection

The data was gathered from the database of the company followed their permission and privacy policies, including:

- Customer Information: name, mobile contact, address.
- Customer demand for each month.
- Quantity of vehicle fleets, their capacity (bottle) and travelling costs (VND).
- Location of warehouse, distribution centers and their capacity (bottle).

3.4. Solving Development

In the literature, many approaches were introduced as appropriate solving methodology for a mathematical programming model, such as Hybrid Ant Colony Optimization Algorithm, Genetic Algorithm, Simulated Annealing, etc. Those approaches are legit and widely applied for optimization problems; however, their instances are mostly in large and mega scale. In the paper, the presented case is categorized as small to medium scale, as the characteristics of business are high-end products and business-to-business trading. Thereby, implementing Commercial Optimization Package (COP) will be considered as a solving approach for this research. Unlike previous approaches, using COP has distinct characteristics since it is developed to solve only some specific optimization issues. Its procedure is practical, easy to implement and often delivers quick and understandable solutions. To save time and effort, several writers choose to use COP after analyzing problems and constructing mathematical models.

Due to the dissimilar nature of the problem statement which is to find a solution for a luxury product distribution network given limited facilities but in medium scale, the most suitable approach should be the advantages of commercial optimization packages and other aiding software, including:

- ***CPLEX Optimization Studio***, or ***CPLEX optimizer***, is the mathematical solving package belonging to a global technology group IBM. Fascinating feature of CPLEX is its flexibility for a wide range of models, from integer programming, complex linear programming, quadratic programming problems to mixed-integer programming problems. Moreover, the solver also has a multi-interface to C++/C#, Python, Java, MATLAB, and Microsoft Excel. Lastly, CPLEX provides a user-friendly language

and simple manipulation which allows the users to input and translate their proposed models easily.
- ***Microsoft Excel*** is a solving tool apart from the Microsoft Office package that specializes in processing data in table form. The software is designed to record, control data, conduct calculation, and build up visual statistics. Moreover, it also allows integration with several add-in software. In this paper, the mentioned applications are Excel VBA and Power Query.

In conclusion, the two applications will oversee the recording data, solving, and displaying results processes for the proposed mathematical programming models and their constraints.

3.4.1. Handling Data Phase

As the rough data above are not compatible with the optimization program formats, rearrangement and calculation procedure are required before the solving phase.

One critical input for every routing problem is the distance matrix, which is basically the road distance between each and every point. Technically, those data could be extracted by entering the departure and arrival point on any map applications such as Google Maps, Apple Maps, Bing Maps, etc. However, for the larger scope problem where the given locations can go as high as hundreds, or thousands, it is impossible to do it manually. Thereby, there is a need for a proper method to calculate the distance matrix regarding the requirements of speed and accuracy.

Generally, the paper will implement two applications:
- Bing Maps REST Services Application Programming Interface (API) – an interface which allows two software to "talk", or exchange information to each other.
- Microsoft Excel and its add-in features, including Visual Basic for Applications (VBA) & Power Query.

The process of forming input data includes two sub-phases:
- Geocoding, a process of converting addresses (like a street address) into geographic coordinates (like latitude and longitude), using Bing Maps API and Power Query.
- Calculate driving distance between given locations, using Bing Maps API and VBA.

Departure Node	3
Arrival Node	4

Departure Coordinate	21.01808,105.81178
Arrival Coordinate	21.00026,105.8031
Key	Agl3VNMxrmVUFhhuousPAEOMA5blPGGU7Ovn5vDGF4kccaJRKkJRR1Hips6cHeg9

Distance	3

User Input

Function Output

Figure 3. Illustration of Geocoding

Raw Data from The Retailer

Address	District	City
Ngõ 95 Chùa Bộc, Trung Liệt	Đống Đa	Hà Nội
125 Thái Hà, Trung Liệt	Đống Đa	Hà Nội

Geocoding Results

FormattedAddress	Latitude	Longitude	Coordinate
Ngõ 95 Phố Chùa Bộc, Hanoi, Hanoi, 11515	21.00621	105.82814	21.00621,105.82814
Phố Thái Hà 125, Hanoi, Hanoi, 11515	21.01294	105.81992	21.01294,105.81992

Figure 4. Illustration of Driving Distance Calculation

3.4.2. Solving Phase

Model 1

Model 1 intervenes the problem from distribution centers to customers and is built based on the concept of multi-depot and multi-trip problem, in which each depot (or distribution center in this case) has its own storage capacity and fleets of vehicles, and each vehicle has its own capacity. Each customer has a demand in one period and must be supplied in the same time horizon. However, as the fleet of vehicles is limited in quantity and various in capacity, they must operate several trips in one period. In addition, each trip is bound with three conditions: capacity of operating vehicle, limitation of traveling distance, and capacity of distribution center that holds the products delivered in that trip. In a specific way, the constraints are developed to assign the number of trips until all customer's demands are satisfied, and for each trip, the model will decide: the customer's ID serving in that trip, the DC holds goods for those customers, and the performing vehicle.

For model 1, the input data includes: (1) Capacity of each DC; (2) Capacity of each vehicle at each DC; (3) Demand of customers for each period; (4) Distance matrix between customers and customers; (5) Distance matrix between DC and customers; (6) Transportation cost per km of each vehicle at each DC.

Accordingly, the model 1 will return the results include:

(1) The transportation cost covers from DCs to customers in each period.

(2) The quantity of trips that one vehicle at each DC must perform in a period. As in figure 5, DC 1 assigns its vehicle 1.1 to perform 2 trips while DC 2 assigns its vehicle 2.1 to perform 1 trip.

(3) Decide the customer sequence for each trip. As in figure 5, in trip #2 of vehicle 1.1, drivers must travel following the order: from DC 1 → customer 2 → customer 9 → customer 7 → customer 4 → back to DC 1.

(4) Decide the cargo load of each vehicle in one trip. As in figure 5, cargo load of vehicle 1.1 is the sum of demand of customer 2, 4, 7, 9.

(5) Decide the replenishment orders of each DC and quantity per each order in one period. As in figure 5, the replenishment order #1 of DC 1 is the sum of the cargo load of all vehicles 1.1 and 1.2; moreover, there will be 2 replenishment orders for DC 1 in that period.

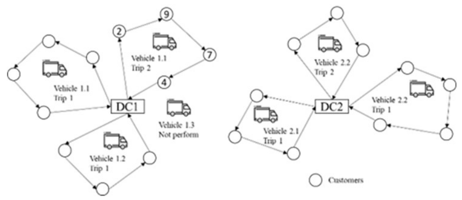


Figure 5. Visualization of Model 1 Results

_________________*Model 2*_________________

The model 2 is built based on the concept of partial shipment and covers the layer from warehouse to distribution centers. In other words, after receiving the replenishment orders from distribution centers, which is one of the outcomes from model 1, model 2 will be executed to assign replenishment trips for vehicles at the warehouse. The general view for this model is that in one period of demand, each DC can have multiple replenishment orders, and the criterion for warehouse is all inbound goods must be shipped to DC before the delivery trips to customers is performed, but simultaneously for each DC. In other words, order #1 from DC 1 and order #1 from DC 2 will be shipped at the same time. Moreover, to prevent the case when the good quantity in one order is larger than the vehicle's capacity, model 2 is designed to split the order and assign those to different vehicles if needed.

For model 2, the input data includes: (1) Capacity of each vehicle at warehouse; (2) Distance matrix between warehouse and DCs; (3) Demand of DC/ Replenishment orders of DC from model 1; (4) Transportation cost per km of each vehicle at warehouse.
Accordingly, the model 2 will return the results include:

(1) The transportation cost covers from warehouse to DCs in each period.
(2) The quantity of replenishment trips that one vehicle must perform in a period. As in figure 6, vehicle W1 must operate two replenishment orders to DC 1 in one period but not overlap with each other. For example, vehicle W1 is chosen to operate replenishment order RT #1 for DC 1; thereby, while considering a vehicle to deliver RT #1 for DC 2, vehicle W1 must be eliminated from the list.
(3) Decide the quantity of vehicles needed for one replenishment order. For example, in figure 6, to fulfill the replenishment order #2, two vehicles W1 and W2 are required.
(4) Decide the cargo load of each vehicle in one trip.

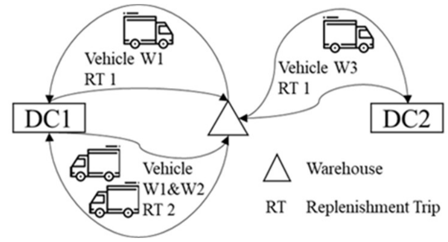


Figure 6. Visualization of Model 2 Results

4. Results and Discussion

The mathematical model was programmed on CPLEX 12.10.0 and ran on Intel ® Core ™ i7-10510U CPU 2.30 GHz with 8.00 GB RAM. The two models were run with different data sets, where specifications are provided in the table below. Moreover, the two models were also run with different periods of time in a year, including: low demand season, normal demand season, and high demand season. The processing time, as well as the objective values were also presented in the table below.

Table 1. Model 1 – Results

	Dataset 1.1		
	Low Demand Season	Normal Demand Season	High Demand Season
Objective value (VND million)	7,4	9,670	12,240
CPU time (mm:ss:ff)	00:05:48	00:07:39	00:08:41

	Dataset 1.2		
	Low Demand Season	Normal Demand Season	High Demand Season
Objective value (VND million)	21,28	22,15	12,240
CPU time (mm:ss:ff)	00:16:98	00:32:22	00:30:08

	Dataset 1.3		
	Low Demand Season	Normal Demand Season	High Demand Season
Objective value (VND million)	2,136	2,2	2,298
CPU time (mm:ss:ff)	00:29:35	00:59:96	02:19:27

	Dataset 1.4		
	Low Demand Season	Normal Demand Season	High Demand Season
Objective value (VND million)	3,093	3,63	4,188
CPU time (mm:ss:ff)	05:10:08	10:37:44	18:42:57

Table 2. Model 2 - Results

	Dataset 2.1		
	Low Demand Season	Normal Demand Season	High Demand Season
Objective value (VND million)	132,744	208,6668	282,096
CPU time (mm:ss:ff)	00:01:26	00:01:37	00:01:38

	Dataset 2.2		
	Low Demand Season	Normal Demand Season	High Demand Season
Objective value (VND million)	72,76	145,444	148,608
CPU time (mm:ss:ff)	00:01:62	00:01:37	00:01:16

In general, the two models can process efficiently in a reasonable computation time, with different data sets of different characteristics.

Table 3. Model 1 – Input Data Details

Data Set	Customer Qty	DC Qty	Vehicles	
			Type	Quantity
Set 1.1	15	1	3.5 tons 2.5 tons	1 1
Set 1.2	33	1	3.5 tons 2.5 tons	1 1
Set 1.3	30	2	*DC #1* 1.5 tons 2.5 tons *DC #2* 1.5 tons 2.5 tons	---------- 1 2 ---------- 1 1
Set 1.4	57	2	*DC #1* 1.5 tons 2.5 tons *DC #2* 1.5 tons 2.5 tons	---------- 1 2 ---------- 1 1

Table 4. Model 2 – Input Data Details

Data Set	Customer Qty	DC Qty	Vehicles	
			Type	Quantity
Set 2.1	90	3	6 tons 5 tons 1.5 tons	1 2 1
Set 2.2	45	3	6 tons 5 tons 1.5 tons	1 2 1

Table 5. Vehicle Capacity

Types	Capacity
1.5 tons	2000
2.5 tons	3330
3.5 tons	4600
5 tons	6660
6 tons	8000

Table 6. Distribution Capacity

DC	Capacity
#1	8000
#2	8000
#3	6000

As mentioned above, model 2 is dependent strongly on the results of model 1, leading to its detailed characteristics being the combinations of those of model 1. To be specific, there are 4 sets of data for model 1, divided into two classifications: North Area (set 1.1 and 1.2) with one distribution center, and South Area (1.3 and 1.4) with two distribution centers. Thereby, the input data for model 2, whose goal is to deliver replenishment orders for all DCs in the nation, is presented with a quantity of three.

Moreover, for each region, small scale data (1.1 and 1.3) were extracted randomly from the whole set and tested first (1.2 and 1.4) for the adaptability testing of the models. The tables have shown that the model worked well with small size data, as the computation times were extremely short. Plus, after the execution phase, no conflicts among the constraints or errors have been found in the application's catalog box. For medium scale data sets (30 and 57 customers), the models were also applicable but generated results in slower time since the variables were marked up to over 100000 units (recorded from the application).

The next characteristic is the pattern of demand and distance. To be particular, the North Area data set has much larger demand than in the South, but the fluctuation among clients is greatly strong, in which a few make purchases with amounts up to thousand, but a few make with quantities beneath hundred. Plus, the distance matrix for this area is also punctuated since some customers are living outside the central or in the nearby provinces that are 200 to 300 kilometers far from Ha Noi. In contrast, customers living in Ho Chi Minh have a centralized distance pattern since the matrix shows only under 50 kilometers data. This explains the reason why the transportation cost of the North and South has such a distinct gap. Moreover, Southern customer's buying patterns are similar to each other, with around 100 to 600 bottles for one order, and the fluctuation among seasons is also insignificant.

For the capacity utilization aspect, the models were also generated with proposed solutions. As can be seen in table 7, almost all trips show an occupation percentage larger than 90%, which is an outstanding performance. However, there were some between 70% to 80%, which showed the normal pattern, as the demand fluctuated. Plus, the models are bound with the criterion that all customers must all be served. Thereby, even if the cargo load is not full, delivery trips are still generated.

Table 7. Detail Result of Data Set 1.4

*Ocpt: Capacity Occupation, calculated by the formula $Occupation = \frac{Cargo\ Load}{Capacity} \times 100\%$.

DC	Vehicle Assignment Details				
	Type	**Trip**	**Cargo Load**	**Capacity**	**Ocpt.**
1	2.5 tons - #1	1	3104	3330	93.21%
		2	2481	3330	74.50%
	2.5 tons - #2	1	2465	3330	74.02%
		2	3266	3330	98.08%
	1.5 tons	1	1890	2000	94.50%
		2	1850	2000	92.50%
2	2.5 tons	1	3218	3330	96.64%
		2	3295	3330	98.95%
		3	3301	3330	99.13%
		4	3326	3330	99.88%
		5	3325	3330	99.85%
		6	3271	3330	98.23%
	1.5 tons	1	1896	2000	94.80%
		2	1777	2000	88.85%
Total	DC #1		15056	17320	86.93%
	DC #2		23409	23980	97.62%

For routing aspect, the two models also satisfied all of requirements. As described above, the replenishment orders for all distribution centers must be operated in the same time horizon, so the vehicle assignments should not be overlapped. In table 8, the criterion was accomplished. Plus, partial shipment policy was also implemented, as for replenishment trip #1 for DC #1, the order was split and operated by two vehicles.

Table 8. Detail Result of Data Set 2.1

*Dest.: destination
**RT: replenishment trip

Dest.	RT	Vehicle	Cargo Load
DC #1	1	3	5343
		4	2000
	2	1	6974
	3	2	3231
	4	2	4885
		4	2000
	5	1	7603
	6	2	3407
DC #2	1	1	7621
	2, 3	none	
	4	1	7435
	5, 6	none	
DC #3	1	2	4995
	2	3	3295
	3	3	3301
	4	3	5222
	5	3	3325
	6	3	3271

In general, the numerical experiences have proved that the two models are committed to compromising the main objective as well as related sub-goals as stated in the previous sections. First, the proposed models provide a delivery pattern which guarantees that all driving routes are optimized and effectively take advantage of the retailer's limited and heterogeneous resources. The finding extends those from previous studies, which mostly incorporate uniform or unconstrained capacity vehicles. Secondly, the improvements noted in this study were the comprehensive assignments of transportations, which prevented the scenario that possibly occurred while working with heterogeneous capacity vehicles: some are repeatedly used, and some are put in hiatus status. Lastly, the research is compatible with distinguished buying patterns in a year, as fluctuated demand is inevitable for any trading businesses, especially for wine retailers which are still amateur and implicate many boundaries.

5. Conclusion and recommendations

5.1. Conclusion

The more the Logistics Industry in Vietnam grows, the more disciplines are born and developed, and meeting obstacles is inevitable. This research has shown the trammels while operating a new high-end product business, in which the operating and administrative systems are still new and immature, leading to a low-efficient and costly distribution network. Fortunately, all challenges defined by Retailer R Company have been solved in this paper. The two proposed Capacitated Vehicle Routing Problems models, considering various constraints whose highlights are: multi-depot, multi-trip, balance in storage capacity, are proved to work well with every scenario in a reasonable computation time, with exact solutions.

One noteworthy improvement out of the literature is that the mathematical model has expanded and tested with two additional indices, which are the period and vehicle identification. This left the opportunity for the writer to cooperate capacity perspective into the existing multi-depot and multi-trip problems. The modified models propose many useful applications, as they were developed to "deal" with existing transportation resources, not to propose an optimal one. Moreover, the models are committed to contribute around 5 - 15 % to the capacity utilization of vehicles, compared with the observation and rough data provided by the company, but the number could be changed based on different types of buying seasons, leaving room for more innovative and modified approaches in the future research.

The purpose of this study is to optimize and declutter delivery trips among business and customers and internal shipments inside business. The fact that when those trips are planned under entrepreneur resources saving and scientific ways, it can help the operators/ drivers improve their key performance index as well as control the waste in resource utilization. Plus, the managers can also easily follow the shipment progress and detect the troubles if any. Finally, planned trips could also cut off redundant trips, which is not only cost effective, but also reduces the amount of fuel emissions to the environment.

5.2. Recommendations

Vehicle Routing Problems remain many potential perspectives for researchers to analyze and develop. This paper has shown and integrated some of those; however, there still is possible for another cooperation. For future work, the paper aims to extend to time windows and customer services areas, in which the models could plan and control service times of each vehicle at each customer location. In addition, as the velocity of vehicles is treated as a constant variable in the paper, future work also endeavors to achieve and investigate in this field, with other traffic coefficients such as traffic lights, broken vehicles, etc. Moreover, the paper also aims to continue to improve its approach as well as current parameters to prepare for the ability and appliance of models if the company desires to expand its business scale and customer targets.

6. Acknowledgement

The authors would like to send gratitude to all lecturers at the School of Industrial Engineering and Management in International University - Vietnam National University for all support and evaluations; to Retailer R Company for the opportunity to access and investigate their business.

References

Ambrosino, Daniela, and Maria Grazia Scutellà. (2005). "Distribution Network Design: New Problems and Related Models." *European Journal of Operational Research* 165(3): 610–24.

Deng, Shuai, Yanhui Li, Hao Guo, and Bailing Liu. (2016). "Solving a Closed-Loop Location-Inventory-Routing Problem with Mixed Quality Defects Returns in E-Commerce by Hybrid Ant Colony Optimization Algorithm." *Discrete Dynamics in Nature and Society* 2016.

Guo, Hao et al. (2018). "A Nonlinear Integer Programming Model for Integrated Location, Inventory, and Routing Decisions in a Closed-Loop Supply Chain." *Complexity* 2018.

Liu, Bailing, Hui Chen, Yanhui Li, and Xiang Liu. (2015). "A Pseudo-Parallel Genetic Algorithm Integrating Simulated Annealing for Stochastic Location-Inventory-Routing Problem with Consideration of Returns in e-Commerce." *Discrete Dynamics in Nature and Society* 2015.

Petch, R. J., and S. Salhi. (2003). "A Multi-Phase Constructive Heuristic for the Vehicle Routing Problem with Multiple Trips." *Discrete Applied Mathematics* 133(1–3): 69–92.

Rahbari, Misagh et al. (2020). "Modeling and Solving a Five-Echelon Location–Inventory–Routing Problem for Red Meat Supply Chain: Case Study in Iran." *Kybernetes*.

Salhi, S., and M. Sari. (1997). "A Multi-Level Composite Heuristic for the Multi-Depot Vehicle Fleet Mix Problem." *European Journal of Operational Research* 103(1): 95–112.

Wassan, Naveed, Niaz Wassan, Gábor Nagy, and Saïd Salhi. (2017). "The Multiple Trip Vehicle Routing Problem with Backhauls: Formulation and a Two-Level Variable Neighbourhood Search." *Computers and Operations Research* 78: 454–67. http://dx.doi.org/10.1016/j.cor.2015.12.017.

Zhen, Lu et al. (2020). "Multi-Depot Multi-Trip Vehicle Routing Problem with Time Windows and Release Dates." *Transportation Research Part E: Logistics and Transportation Review* 135(January): 101866. https://doi.org/10.1016/j.tre.2020.101866.

REAL ESTATE BUSINESS IN THE FACE OF CHANGES IN CUSTOMER BEHAVIORAL PSYCHOLOGY DURING THE COVID-19 PANDEMIC IN VIETNAM

***Nguyen Minh Ngoc*[1], *Nguyen Thi Hai Yen*[2], *Nguyen Hoang Tien*[3]**

[1] *University of Finance – Marketing, Ho Chi Minh city, nmngoc@ufm.edu.vn, 0912359472*

[2]*National Economic University, Hanoi, haiyen3579@gmail.com*

[3]*University of Food Industry, Ho Chi Minh city, vietnameu@gmail.com*

Abstract: This study aims to present opportunities and challenges for Vietnamese real estate enterprises about the change in consumer behavioral psychology in the post-Covid-19 era. The research has selected a number of researches related to consumer behavioral psychology that can collect analytical data, predict the situation and give specific factors for the research. The research results show that in the post-Covid-19 period, real estate businesses can still seize many opportunities for them to return to normal. Especially, through the covid-19 period, real estate businesses restructured, set up projects in different provinces, saved costs, reformed core personnel and developed new real estate products to change customers' habits and behavior of real estate consumption. New sales mindsets are also formed and developed today when the epidemic in the world is still going complicated. This is also the period when businesses can make better policies, with better plans for the upcoming plans. Which real estate businesses really stay strong in the real estate industry and seize opportunities to bring their businesses out of this crisis. However, the challenges ahead are huge when Vietnam is experiencing the 3rd wave of covid-19, the strongest wave spreading in 45/63 provinces and cities nationwide. The worrying thing about the business is its finance, personnel, especially the consumer behavioral psychology of customers after Covid-19 has changed, they choose greener products, fear of risks than before. From the perspectives on opportunities and challenges, the research has also given a number of solutions that can help real estate businesses refer and give directions for businesses in this period.

Keywords: Real estate business, consumer behavioral psychology, Covid-19 pandemic, logistic real estate.

1. Introduction

Real estate consumers have more and more choices, more opportunities to find themselves a suitable product at an affordable price. An in-depth understanding of real estate consumer behavioral psychology is the key to a successful marketing strategy both domestically and internationally. This is one of the first important reasons that attracted us to study and research on this issue. In the face of the current complicated epidemic situation, people's consumption and investment behavior in real estate has changed quite a lot. Consumer behavior is a field of research that has its roots in sciences such as psychology, sociology, humanities, and economics. Understanding consumer behavioral psychology will provide the foundation for marketing strategies, such as product prediction, market segmentation, new product development, new market applications, global marketing, and marketing mix, ...

The above are some of the main reasons why we choose the post-Covid-19 consumer behavioral psychology in Vietnam to explore and research. Our research objectives are first of all to survey the real situation of customers, the situation of changing consumer behavior. We believe that this research should be done regularly to be able to monitor the rapid fluctuations of the real estate market in Vietnam, in order to offer timely and appropriate solutions.

Stemming from that awareness and realizing the changing of consumption behavior of people, especially in the current epidemic context. Therefore, we have chosen the topic "Real estate business in the face of changes in customer behavioral psychology during the Covid-19 pandemic".

2. Theoretical framework

2.1. Factors impacting consumption behavioral psychology

Consumer behavioral psychology is the act of a person buying and using a product or service, including both psychological and social processes occurring before and after the act. The study of consumer behavior includes the study of the individual consumers, how they choose products and services and the impact of this process on consumers themselves and society.

There are four psychological factors that influence consumer behavior: Motivation, perception, learning, and attitude or belief system. Motivation speaks to the internal needs of the consumer. Understanding how to motivate your customer is a powerful tool. The way your target customer perceives the world or learns about your product, whether online or in person, can also influence behavior. Finally, belief systems have the ability to influence all of the above. For example, some people learn to best visually. Professional pictures and images of your product or service can communicate a thousand words regardless of belief system. This explains why pictures and images are so important for marketing.

Consumers make purchase decisions every day. When they make a decision about your product, they're thinking about solving a need. That need may be driven by a variety of factors. Each of the factors listed above can be related to ways in which a real estate business can utilize its available resources to increase the probability of a sale.

2.2. Main types of behavioral psychology of real estate consumer in volatile market

In the context of the sluggish economy, the real estate market was also greatly affected. Typical behavioral psychology models of customers in the real estate market during a difficult period can take place as follows (Yen N.T.H., 2016):

Availability Bias

A form of deviation due to the heuristic effect. It is the psychology of recalling the rules that are available in memory. Deviation from established rules can lead to overestimation of the probabilities of overt or critical events, especially recent occurrences.

Situational bias causes consumers to buy and sell based on available information (advertisements, recommendations from consultants, friends...). Consumers spend based on the list of information available in memory. In their mind, information other than the list is not easily elicited and overshadowed. Customers who fall prey to situational bias will choose to buy properties that are limited to their knowledge. Adair et al. (1994) show that consumers frequently pour money into information-rich projects. Market imperfections related to the availability of information can cause investors to deviate from the normative process. Availability can also lead to representativeness bias: consumers often mistakenly assume that price movements over the current time period are predictive of the overall distribution of benefits profit. Ling (2005) also proposes a similar model for professional forecasters in the commercial real estate market.

Mental Accounting

People sometimes tend to make individual decisions that should be combined. For example, many people have divided their family budget into two parts, one for food and the other for entertainment. Sentimentality means that consumers use experience to divide the factors of consumption into different affective calculations (Thaler, 1985). Almenberg and Karapetyan (2009) studied sentiment calculation in the housing market and found that it makes capital structure inefficient. In general, households use a near-optimal debt structure to finance their home. Campbell (2006) finds that a significant portion of households' holdings produce serious investment errors. The biggest mistakes include: not participating in risky asset markets, not diversifying portfolios, and failing to implement collateral refinancing methods. In general, the poor and the less educated are more likely to make these mistakes than the rich and educated (Campbell, 2006).

Loss aversion

A psychological state in which people fear loss, the psychological impact of loss is greater than the impact of an equivalent gain. The fear of loss is also gradually becoming popular with investors in the market. This psychology can be explained on the basis of prospect theory, developed by Tversky and Kahneman (1979). Deviation due to fear of loss will make customers stick to low-priced real estate because they have the mentality of waiting for prices to rise again before conducting transactions. However, because of this indecision will push them into heavier losses. In contrast, investors with this psychological bias often easily sell properties with an uptrend because they are afraid that real estate prices will quickly decline. At that time, most customers thought they had greatly reduced their risk, but in fact the risk increased when the portfolio became unbalanced.

The fear of loss causes real estate owners to set a higher selling price or limit price (minimum acceptable price) when the value of their property has decreased since the original purchase, which can be seen as a way to avoid a loss equivalent to the purchase price. The effect of loss aversion is that it can prevent real estate transaction prices from exhibiting a decline during the early stages of a market downturn (Sheharyar, B. & David, G., 2010). These authors also show that experienced customers and sophisticated institutional investors are both more sensitive to loss than individual investors and retail investment firms.

Herd behavior

According to Keynes (1936, 1937), herd behavior reflects people's responses to uncertain problems and limited information. When faced with uncertain problems, people are more inclined to consider the behavior of others, because the majority may have more accurate information or because there is assurance in the quantity. In other words mimicking each other. Some of Keynes' reliable findings on common behavior are recognized by experts: in volatile markets, the real estate investor does not just follow his or her best judgment about direction future property value but also consider the judgments of others. Because they are well aware that they cannot fully update all the information, so it is wiser to refer to the action as well of others before making decisions for yourself. These views are being developed in modern herd behaviorism, which focuses on rational motivations to ignore personal information when the actions of others provide valuable information (Banerjee, 1991; Bikhchandani et al., 1992; Chamley, 2004; and Scharfstein & Stein, 1990).

“Hand behavior – people act according to what others are doing, even when their own information suggests they should act differently” (Banerjee, 1992). For single consumers, because of the limitations in owning the information as well as the considerations mentioned above, these people are easily caught up in the “trending games” or Simply put, the "price-making" games of organizations. At the same time, rumors, low-quality information out-of-the-box are sometimes "fully utilized" by consumers and end up creating a crowd that acts the same in an unusual way reasonable.

Herd behavior is considered a cause of mispricing and speculative bubbles (Shiller, 2005). Shiller (2006) argues that the housing boom should be viewed as a social pandemic. The extent to which speculators follow the crowd is highly dependent on actual profits (Lux, 1995). This supports the view that serial correlation is an important factor in the housing market.

Confirmation bias

Self-confirmation bias or pro-information bias is a state of mind where consumers value supporting information and do not value information that contradicts consumer's predictions. Farlow (2004b) recognizes the presence of self-confirmation bias in the real estate market. He observed that the people who succeeded during the real estate boom were mainly due to wise investments. In contrast, failed investments that are attributed to bad luck, to the influence of others, or to other factors such as the market, or to a decrease in the number of investors - are all reasons for worsening consumer sentiment. A similar form of bias affects future expectations. In general, people tend not to easily adjust their expectations because they are always looking around, looking for a logic that explains and reinforces their beliefs.

Momentum effect

Another type of bias occurs when movements in property prices are viewed as a basis for future price expectations. This response to price changes is often referred to as the “price inertia effect”. Case et al. (2003) investigate whether rising prices encourage people to buy more homes. On average, more than 80% of respondents confirmed that rising prices stimulated them to buy a home. These findings confirm apathy towards fundamentals and the presence of price inertia. Real estate agencies ignore the law of probability and instead overreact to trending news. Another possibility is the presence of herd behavior. Lux (1995) provides a formal model that explains bubbles by linking price inertia and herd behavior. He explained the real estate bubble was a systematic infection process among consumers. This causes the equilibrium price to deviate from the fundamentals. He also emphasized the willingness of consumers to follow the crowd and the importance of realistic returns on future prospects, thereby alluding to price inertia in real estate consumptions.

Money illusion

Money illusion is one of the most common market anomalies in economics. Money illusion in real estate refers to the mistake consumer make when valuing alternatives during periods of inflation due to the difference in nominal and real values. It plays an important role in real estate because it is often associated with periods of inflation in the whole economy. Farlow (2004a) adds that very few people relate inflation to the housing debt they have to pay. If inflation is low, they forget that they are delaying the repayment of their debt that should have been done at a faster rate. Brunnermeier and Julliard (2008) investigated whether money illusion could explain market inefficiencies. And the results indicate that the mispricing is partly due to inflation. Shiller (2008) also confirms and interprets these results. Because homes are an important purchase and sale, we often remember the price of a house from a long time ago. Therefore, the contrast between today's property prices and the previous ones is of course more noticeable than the change in the price of a loaf of bread over that period. This leads to the misconception that a home is a "grand" investment, when the real returns over the long term are very limited (Eichholtz, 1997). Ackert et al. (2011) studied the relationship between money illusion and price expectations using survey data. They scoured the literature for homeowners who had been perceived as suffering from money hallucinations, but whose price expectations were reasonable. Hayunga and Lung (2011) provide empirical evidence that although money hallucinations are a persistent feature of homeowners, overconfidence has more of an impact on mispricing.

Home bias

Traditional home country bias means that individual consumers invest more in the housing market in their home country, as explained by Karlsson and Norden (2007). Goetzmann (1993) empirically studied real estate diversity and found that regional diversification prevails over real estate diversification in a locality. The risk frequency of owning four homes in many different areas will be lower than the risk frequency of owning thousands of homes in just one area. Farlow (2004b) confirms that most investors are too attached to just one asset - their home. Most decisions are not related to marginal trading interests. The explanation is: the risks of owning a home are low.

3. Research methodology

The methods that the article uses to collect data: Search secondary data from newspapers, magazines, the website of the General Statistics Office, organize market research, filter documents about research books fast moving consumer goods. Methods of analysis and theoretical synthesis: Analysis is the study of different documents and theories by analyzing them into parts to better understand the object. Summary is a link on each side, each piece of information has been analyzed to create a new complete and insightful theoretical system of the object.

Classification method and theoretical systematization: Classification is to arrange scientific documents on each side, each unit, each problem has the same natural sign, the same direction of knowledge development into a system based on the theoretical model to understand the object more fully.

Data analysis method: used to describe and analyze data to draw conclusions, so that the analysis results can be effectively compared with the actual data of real estate assets to adjust policies of real estate businesses.

4. Research results and discussion

4.1. The change in behavioral psychology of real estate consumption in the period of Covid-19 in Vietnam

In both 2020 and the first quarter of 2021, according to the authors' compilation and analysis of documents, the real estate market has a decrease in size but not significantly. Vietnam's GDP growth is relatively positive during the global health crisis. The Vietnam's list of feats this past year has been remarkable given the current health and economic hardships across the world. Vietnam had incredible success containing Covid-19, with the country of 96 million people registering only 2,631 cases and 35 deaths at time of researching. During the same period, Vietnam eclipsed all of its Asian countries economically by posting GDP growth of 2.9%, a noteworthy feat considering the GDPs of neighboring Thailand and Malaysia contracted by 6.1% and 5.6%, respectively, in 2020. Consequently, Vietnam's real estate market has developed in 2020, with continued economic growth leading to a surge in property prices. The country's real estate market proved resilient during the pandemic, with both the industrial and residential sectors leading the pack.

As early as May 2020, quarantine restrictions were lifted and citizens were able to resume everyday life, with the country essentially operating a fully functioning internal economy while the rest of the world remained in various forms of lock down. As factories reopened and construction projects resumed, Vietnam's real estate market recommenced its record run (Philip, H., 2021).

Logistics real estate

Vietnam is one of the countries with a stable, developing economy and a large number of young people. During the epidemic, Vietnam was and is currently efficiently controlling the disease, assisting Vietnam in becoming the focus of many international firms. Due to this, Vietnam's industrial and logistics real estate has become one of the main manufacturing and logistics destinations for investors. The rise of the retail and e-commerce industries, particularly during this time, has boosted the corporate need for manufacturing, logistics real estate and storage throughout Vietnam.

The development of e-commerce means that 'last mile' fulfilment has tremendous potential here, and demand will increase for well-located warehouses on central business district boundaries and on important city arterial routes.

Due to rising demand, 2021 will be a chance for new industrial clusters. Many provinces and towns that were not historically industrial centers are now increasing investment in infrastructure, industry, and warehouses to keep up with the new wave of an industry that has exploded globally over the last 25 years. The Vietnamese government is constantly monitoring the supply of new industrial land and warehousing, and if there's a potential lack of supply, they tend to approve new locations and investments quickly.

Vietnam's two biggest cities are Hanoi and Ho Chi Minh City. The northern capital of Hanoi provides advantageous logistics real estate opportunities. The most populous city, Ho Chi Minh City, is in the south and is recognized as Vietnam's logistics real estate powerhouse.

These reasons have created a very positive sentiment for foreign and Vietnamese logistics real estate investors. This not only affects the psychological behavior of large investors, but also affects the psychology of retail investors, who invest in logistics services (Tien, N.H., Ngoc, N.M., 2019). Worldwide warehousing behemoths such as GLP, LOGOS, and JD.com have been actively engaged and invested in industrial and logistics real estate in both the North and the South since 2020. Vin Group, a large Vietnamese real estate developer, has also joined the market, with two new industrial parks planned to accelerate the implementation of the first industrial real estate projects in 2021.

The fast growth in foreign direct investment capital, together with the value chain change, has paved the way for a future for Vietnam's logistics real estate market. With the relocation of manufacturing facilities from China to Vietnam and the fast expansion of local demand, logistics real estate in Vietnam has been the top choice for foreign investment, in particular, the logistics real estate market.

Residential

On to Vietnam's housing market, which has also seen unprecedented growth in recent years. As local Vietnamese have limited investment options outside of the housing market, rampant demand for apartments has exceeded the supply of units, with many new developments selling out shortly after sales launch. According to Cushman & Wakefield, apartment prices in HCMC have risen in response and grew by a staggering 90% in three years from 2017 to 2020, including by 12.8% in 2020 alone. While demand from foreign investors is one of the factors pushing the market, majority of the growth comes from Vietnamese citizens (in any new apartment development, foreigners are only allowed to own up to 30% of the total units). As such, the country's economic progress coupled with a rapidly expanding middle class are the primary demand drivers for the expanding housing market.

However, there are some pretty clear signs of an imminent decline. However, there are some pretty clear signs that an impending decline as defensive and fear of loss is starting to dominate the market in Q1 2021. This suggests a change in sentiment. Consumer behavior management is quite fast and sensitive in the context of the ongoing pandemic and increasingly risky trends.

Office

According to Paul Fisher, country head of global real estate firm JLL in Vietnam, "Vietnam will continue to use a priority office model. The lack of face-to-face interaction has put pressure on properties team, and while some customers look forward to adopting flexible working practices in the future, much of this will include the office that remains the central point for business activities". As the future of the office remains uncertain worldwide, working habits in Vietnam are unlikely to change due to the country's short closure period as well as limited technology infrastructure.

According to Savills Vietnam, office projects for lease in Ho Chi Minh City have sustainable growth even during the Covid-19 pandemic. The occupancy rate has always been at 96% and office rents in the first quarter of 2021 will continue to increase from 2 to 4% compared to the same period of 2020.

However, there are some companies with low budgets that have moved to lease Grade A office space to lease Grade B office space in response to the pandemic that may still be ongoing.

Hotels

Like the rest of the world, 2020 wreaked havoc on Vietnam's hotel market, with occupancy hovering between 20-30% for most of the year and first quater of 2021. While the recovery is expected to be gradual, the outlook remains strong given the country's travel industry was taking off pre-pandemic. International arrivals grew from 3.8 million in 2009 to over 18 million in 2019, as economic progress led to increased business travel while Vietnam simultaneously became a sought-after vacation destination for foreigners. Vietnam's government is encouraging this growth by heavily investing in travel infrastructure: construction has started on a second airport in HCMC with a total investment of $4.7 billion, while Da Nang opened a new international terminal in 2017. More recently, Hanoi's Department of Planning and Architecture recommended the city start preparing to build a second airport to accommodate increased travel demand in the coming decade. With the government targeting 6.5% GDP growth in 2021, Vietnam's real estate market is poised to ride the economy's tailwinds into the future.

Despite the positive macro information, the hotel market continued to be negative. From mid-2020, the vast majority of investors refuse to buy hotels from 1 to 3 stars, even 4 stars. The typical psychological and behavioral states of hotel investors at this stage are momentum effect and fear of loss when hotel prices go down and revenue plummets. At the same time, herd behavior is also evident in the hotel market.

Subdivision land plot

The trend of splitting land plots to sell to real estate speculators has taken place very strongly in the last 5 years in Vietnam. This trend does not decline but tends to increase strongly in 2020 and the first quarter of 2021. This situation persists and occurs throughout Vietnam to the extent that the government has to issue many directive documents to limit it. The reason for this situation is that land is the most preferred commodity by Vietnamese people to invest in real estate, associated with the people's long history of farming. Besides, the fact shows that for nearly 40 years, holding land plots has always brought great profits to the holders. Moreover, due to the large amount of money supply from the State in 2020 and the amount of remittances that did not decline during the Covid-19 pandemic (actually $ 6.1 billion in Ho Chi Minh City while it is expected to 5.5 billion). So a large amount of money has been invested in small plots of land, which are products of small and medium-sized companies subdividing. The price of these lands usually has a very strong increase of 30 to 50% within 1 year. However, not all speculators and investors are successful.

Although the covid-19 pandemic takes place around the world and Vietnam may pose similar risks to the economy and people's health, the research results show that the more the epidemic occurs and lasts, the more speculators and investors invest in subdivided land plots. It was only when Vietnam implemented a wide and drastic social isolation that this movement temporarily subsided. However, it is certain that this phenomenon will explode again after the pandemic ends.

4.2. Sustainability of new behavioral psychology in real estate consumption in Vietnam

The COVID-19 pandemic has had a strong impact on real estate consumption and investment in Vietnam. The middle class in the market often reacts quickly when turning their attention to safe, low-density real estate products and integrating technologies to prevent future epidemics. Therefore, this trend will be sustainable and continue to be sought in the medium and long term in the future. In addition, the rich are more demanding with the increasing green standards of real estate, both psychologically satisfying, and reducing or eliminating stress during the isolation period. This trend of demand will also take place sustainable in the future.

The pandemic will also forever change the attitudes of families about consumption and property ownership, pushing them to greener consumption and the suburbs or near the forest and sea areas will be their destination. Therefore, the trend of increasing land prices in areas with natural conditions suitable for the above demand will increase prices. The level of price increase can be natural according to the rules of supply and demand in the market or depends on the active influence on consumer psychology.

Besides, it is necessary to confirm that short-term speculative activities, mainly land plots divided into lots, will gradually eliminate non-professional investors due to the fear of loss and the lack of transparency of the market. Therefore, the ability to increase price or liquidity of subdivision land will depend on the level of population concentration of that area due to home bias.

4.3. Opportunities and challenges for Vietnamese enterprises

Opportunities

Despite being affected by the epidemic, this is also an opportunity for businesses to renew themselves, find new directions in accordance with the market needs to survive and grow stronger, because the risks are always along with the opportunity. Sensitive real estate businesses will change the state from "freezing" to immediately seizing new opportunities for development. This is the basis to promote capacity building of the economy, as a premise for a more sustainable development.

Currently, Vietnam's prestige and position are highly appreciated by the international community through the successes achieved from the Covid-19 epidemic prevention last. This is a "golden" opportunity for the world to know about Vietnam with a special advantage of "strategic trust", a safe real estate investment destination and ready to welcome capital flows transferred to Vietnam. In the context that the Covid-19 pandemic is still having complicated developments in many countries around the world, the good control of the pandemic in Vietnam has created a great advantage to be able to be one step ahead in tourism and resort real estate recovery.

The impact of Covid-19 translation creates new awareness, a shifting trend in behavioral psychology consumer, new way transactions of real estate on the principle of distance, limited contact; emerging new online services lines based on digital economy, creating changes in demand for some essential products and real estate services such as digital transformation, e-logistics; providing new market opportunities for Vietnamese enterprises to take advantage of to restructure product of real estate, form new value chains and develop breakthroughs. This is an opportunity for Vietnamese businesses to recognize and re-examine their true capacity, resilience, adaptation to market events, non-traditional security; review, evaluate, restructure real estate resources, customers and find new directions to be more proactive, new business strategies to quickly grasp signals and demand in real estate market. This time is also an opportunity for Vietnamese businesses to focus on renovating equipment and technology at cheaper costs; at the same time, promoting linkages, domestic supply networks, improving quality, reducing product and service prices to take advantage of opportunities to dominate and consolidate market share. At the same time, take advantage a number of international integration policies such as newly signed and officially effective free trade agreements will create advantages for Vietnamese enterprises to participate in the global value chain.

Challenges

The increasing pressure on competition and purification combined with the economic shock caused by Covid-19 has caused many real estate businesses to "wobble". Real estate enterprises in groups of industries directly affected such as hotel, restaurants, education building, play ground, amusement parks ... have fallen into a state of prolonged "hibernation". Real estate projects stalled due to direct impacts from the pandemic and delays in state approvals. The revenue of real estate businesses in 2020 and the year 2021 forecast for the whole year is expected to decrease, even some businesses are losing money or going bankrupt. The distribution of real estate products has been delayed due to the prolonged quarantine in two largest cities, affecting business operations in the following quarters and years.

The sharp decline in people's income has also had a long-term impact on the liquidity of real estate and the financial balance of real estate businesses.

5. Conclusion and Recommendations

5.1. Conclusions

In efforts to fight the global Covid 19 epidemic, the Government of Vietnam is receiving international appreciation for its effective anti-epidemic measures. This is considered by analysts to help reduce pressure and create great advantages for the economy when it reopens after the epidemic.

Researching consumer behavior during the Covid-19 period helps real estate businesses implement marketing campaigns more effectively. The importance of underlying psychological and emotional behavior in decisions to intervene in the real estate market, whether as a consumer or as an investor, is undeniable. Therefore, the content of the article has contributed to further clarifying and promoting real estate businesses to further pursue research on these psychological behavior changes, especially during the covid-19 pandemic and after the pandemic. It also helps real estate businesses understand themselves better so that they can make appropriate adjustments when participating in the real estate market. At the same time, it also helps policy-makers "drive" and regulate the real estate market more effectively. This is especially meaningful for a newly developed real estate market like Vietnam - a market with many fluctuations and ups and downs in recent years.

The results of the study show that real estate businesses need to do the following:

- Pay special attention to market capture and forecast activities
- Fundamentally changing the approach and development in the real estate market with new products, catching up with rapid changes, such as logistics real estate investment. There needs to be rapid evolution like the Covid-19 virus.
- Accept to erase old models, ways and real estate to remake in the new period after the pandemic. Keeping up with changing customer behavior.

5.2. Recommendation

Real estate business have to change to find new markets during and after Covid 19 to match consumers' behavioral psychology in Vietnam, which is an essential task today. Specifically: promoting online sales channels (online), social networks to reach customers according to the capacity of the enterprise; Adjust Media channels based on consumer sentiment.

Actively communicate inside and outside, focusing on advertising to sell. Adjust your media budget, focus on Content Marketing.

Quick response, constant updating of information on the situation of disease change Covid 19 to join hands with the community to show the concern of the enterprise to social issues in order to build accountability and bring the image of businesses come to consumers.

It can be seen that, in the past few days, consumers have changed drastically in their consumption habits, from careful selection of real estate products to buying methods, in general it all comes from want to protect my own health and be highly predictive for the future.

In order to follow this trend, real estate businesses are required to develop new suitable real estate products that meet the growing demand for health and convenience associated with real estate, but with high standards of quality and hygiene for customers willing to pay more.

For distribution businesses, real estate brokers and independent brokers, it is necessary to invest in online channels, develop offline services, change the way customers approach customers and introduce products.

References

Adair, A.S., Berry, J.N., & McGreal, W.S., (1994) Investment Decision Making: A Behavioural Perspective. *Journal of Property Finance*. **5**, 32-42.

Amato, Louis H., & Ronald, P. Wilder., (2004) Global competition and global markets: some empirical results. *International Business Review*. **13**, 401-416.

Brunnermeier, M.K., & Julliard, C., (2008) Money Illusion and Housing Frenzies. *Review of Financial Studies*. **21**, 135-180.

Campbell, J., (2006) Household Finance. Journal of Finance. **61**, 1553-1604.

Case, K.E., Quigley, J.M., & Shiller, R.J., (2003) Home-buyers, Housing and the Macroeconomy. *Paper presented at the Conference on Asset Prices and Monetary Policy*.

Don B., Wendell Mc. Cullock, (1999) International Business - Irwin. Mc Graw Hill.

Eichholtz, P.A, (1997) Long Run House Price Index: The Herengracht Index, 16281973. *Real Estate Economics*. **25**, 175-192.

Farlow, A., (2004a) UK house prices: a critical assessment. *Paper presented at the Credit Suisse First Boston Housing Market Conference*.

Farlow, A., (2004b) The UK Housing Market: Bubbles and Buyers. *Paper presented at the Credit Suisse First Boston Housing Market Conference*.

Goetzmann, W.N., (1993) The single family home in the investment portfolio. *The Journal of Real Estate Finance and Economics*. **6**, 201-222.

Karlsson, A., & Nordn, L., (2007) Home sweet home: Home bias and international diversification among individual investors. *Journal of Banking & Finance*. **31**, 317-333.

Lux, T. (1995) Herd Behaviour, Bubbles and Crashes. *The Economic Journal*. **105**, 881-896.

Ngoc, N.M, Tien, N.H., (2021) Opportunities and challenges for real estate brokers in post Covid-19 period. *International Journal of Multidisciplinary Research and Growth Evaluation*. **1**, 81-90.

Philip Heller (2021) Vietnam's real estate market continues record run. Forbes.com.

Slater, D., (1997) Consumer Culture and, Studies in Advertising and Consumption. Psychology Press. **51**.

Shiller, R.J., (2008) The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It. Princeton: Princeton University Press.

Shiller, R.J. (2006) Long-Term Perspectives on the Current Boom in Home Prices. The Economists' Voice, **3**.

Shiller, R.J. (2005) Irrational Exuberance (2nd ed.): Princeton University Press.

Tien, N.H., Viet, P.Q., Ngoc, N.M., & Anh, D.B.H., (2020) Contemporary Security and Sustainable Issues. Eliva Press, Chisinau, Moldova. 45-46

Tien, N.H., Ngoc, N.M., (2019) Related and Non-related Diversification Strategy of Domestic Business Groups in Vietnam. *International Journal of Advanced Research in Engineering and Management.* **5,** 12-17.

Tien, N.H., Ngoc, N.M, Anh, D.B.H., (2021) Change of consumer behavior in the post Covid-19 period. *International Journal of Multidisciplinary Research and Growth Evaluation*. **2**, 46-52.

Tien, N.H., Anh, D.B.H., & Ngoc, N.M., (2019) Corporate financial performance due to sustainable development in Vietnam. *Corporate Social Responsibility and Environmental Management*. **27**, 695-705.

Tien, N.H., Anh, D.B.H, Ngoc, N.M., & Nhi, D.T.Y., (2019) Sustainable social entrepreneurship in Vietnam. *International Journal of Entrepreneurship*. **23**, 01-12.

Thaler, R., (1985) Mental Accounting and Consumer Choice. Marketing Science. **4**, 199-214.

Yen, N.T.H., (2016) Accessing behavioral financial theory in the above investment research Vietnam real estate market. National conference*: Real estate business – opportunities and challenges in the recovery momentum of the market*. 63-70.

MATHEMATICAL PROGRAMMING FORMULATION

Ngo Hoang Minh Anh[1], Anh Duong[2], Huy Nguyen[3] and Ly Tran[4]

[1] *School of Industrial Engineering and Management, The International University - Vietnam National University, Vietnam, minhanhngohoang@gmail.com*

[2] *School of Industrial Engineering and Management, The International University - Vietnam National University, Vietnam, dvnanh@hcmiu.edu.vn*

[3] *School of Industrial Engineering and Management, The International University - Vietnam National University, Vietnam, nhhuy@hcmiu.edu.vn*

[4] *School of Industrial Engineering and Management, The International University - Vietnam National University, Vietnam, tvly@hcmiu.edu.vn*

Abstract. The demand for Thai Binh Shoes Company's products has increased every year, causing the inventory of work-in-progress to increase. This issue affects safety, increases costs and threatens customer service levels. The thesis analyzed the root cause of this problem and suggested improvements to the system, time study, observation and interview. The final products include an implementation plan for the team's recommendations, a pull system design using CONWIP, and a quantitative impact assessment. Paper research in flowshop production system with collected data to check the results at TBS Company. The CONWIP model considers machine groups and calculates how strictly the number of containers is loaded onto the production line. A good solution was obtained using the program Cplex Optimizer to solve the actual scheduling problem to get the optimal solution. The results show that the amount of inventory is reduced to zero after the end of the horizontal planning period, and the work order is arranged to ensure the optimum between cost and time.

Keywords: Production planning; CONWIP system; Push and Pull

1. Introduction

1.1. Background

Over the past decade, Vietnam's industrial production sector has experienced strong development and has begun to integrate into regional and international markets. State policies also contribute to promoting the development of industrial production. Many industrial parks were formed, many factories were newly invested as well as production was expanded, creating momentum for the country's economic development towards industrialization. This is also an opportunity as well as a challenge for investors and managers in Vietnam.

Manufacturing companies are active participants in the operation of the supply chain, and of course contribute a large part to the overall operating costs of the system. Managers and investors have paid much attention to the operation of their company's supply system, especially the work related to production and inventory that accounts for most of the overall operating costs of the business. In today's tough competitive environment, businesses are facing a lot of pressure from producing a variety of products, improving quality, and reducing production and inventory costs. These are really big obstacles for managers at industrial manufacturing companies in Vietnam. The above difficulties can be solved if the manager has a suitable production plan, the resources are planned and used appropriately.

In fact, the issue of moderation is very important and is widely applied in many different areas in manufacturing and services. Moreover, a clear, specific, easy-to-implement and quickly updated production plan is really necessary for production enterprises.

Thai Binh Shoe Company is a large sports shoe manufacturing company with two main product lines: sports shoes and bags. Currently, TBS Company is facing 3 problems in production:

- Overproduction and underproduction creating a large amount of inventory or loss of revenue are the main problems of the company.
- The number of workers assembling machines may not be enough to meet demand, putting pressure on overtime to meet customer demand.
- The number of assembly and machine workers coming to the workshop outweighs the workload. As a result, there will be many redundant assembly workers, which affects the company's use of time and resources.

To address the above issues, this study will provide:

- It is necessary to develop a plan that requires arranging the appropriate production schedule according to customer needs and operating capacity of the factory.
- Minimizing inventory and penalties for overdue delivery.
- Make recommendations on the optimal number of containers on the conveyor belt to avoid waste or insufficient demand.
- Improve production efficiency and reduce total production cost including holding cost, backorder cost, overtime cost based on arranging production schedule on backlog list reasonably.

The expected output is that the company can easily manage the operation and production processes of the plant. This specific study offers a solution based on the CONWIP system model that helps production managers and executives to make quick and reasonable decisions, especially in adjusting the updated production plan. according to the delivery schedule.

2. Literature review

2.1. Work in process inventory (WIP)

Since pull systems control WIP and observe throughput, their performance is critically dependent on the choice of WIP level. In systems that use cards (kanbans) to govern WIP, setting WIP is done by choosing card counts. To meet customer requirements, WIP levels (or card counts) must be large enough to achieve the desired throughput. They must also be small enough to prevent excessive WIP. Therefore, a basic problem facing pull systems is determining the minimum WIP level to attain desired throughput rate.

Philipoom et al. (1987) used simulation to determine lead times at the workstations and thereby the number of kanbans required at each station to prevent backorders in a dynamic production environment. They also described factors that influence the number of kanbans required in implementing JIT production techniques. These include throughput velocity, process variation, machine utilization and autocorrelation of processing times.

Queueing theory has been used by several researchers to determine the number of kanbans required in stochastic production systems. Bard and Golany (1991) used a mixed integer linear program to determine the optimal kanban policy for each workstation of a multiproduct, multistage production system.

2.2. Constant work in process (CONWIP)

Constants work in process (CONWIP) is a consistent floor flushing technique. The cards direct the progress of the work but are not a clear number; instead of they are the clear work count remaining with an item or cluster over the entire length of the interaction, making it a more sensible strategy when there is a high categorization.

Fowler et al. (2002) believe that Kanban is arranged throughput control while CONWIP is generally more focused on WIP.

In any case, CONWIP can provide more outstanding throughput than Kanban (Spearman and Zazanis 1992). In contrasting CONWIP and push options, Spearman et al. and Hopp pointed out that the confidentiality of CONWIP is considered detrimental to the slightly higher normality of WIP, since WIP in its sporadic push framework and its exhibit collapses immediately when permissions control is reduced. Besides, Herer and Masin (1997) has been proposed a mathematical programming formulation that solves the problem of the order in which the parts are to be produced and explain the major advantages of CONWIP over MRP systems. The objective function of their model is to minimize inventory related costs including holding cost of finished goods and WIP, backorder cost, overtime cost. Optimizing job order and schedule based on mean throughput and flow time using forecasting method is mean value analysis.

Statistical Throughput Control (STC) (Hopp and Roof 1998) is used to vary the number of tokens and to direct the WIP (or throughput) level. However, this requires precise input, which is hard to come by in a complex assembly environment where it can also lead to too many tag changes. According to CONWIP, it is necessary to standardize some items, because assuming that the number of tokens to manage the WIP level, the settlement responsibility of each tag should be equal. Gaury et al. (2000) argues that a weakness of CONWIP is that the inventory within the framework is not independently controlled; High inventories can appear before the machine is underperforming and when the machine splits. Several attempts have been made to tailor CONWIP to the specific requirements of the semiconductor industry. A nonlinear mixed integer programming model for a CONWIP-based production system was developed by Cao and Chen (2005) which the parallel fabrication lines fed an assembly station. In the model, the optimal job sequence and lot sizes were determined, while number of containers or WIP level were not discussed. Luh et al (2000) developed a mathematical programming model for a single CONWIP based on serial production line in the job shop for the Aircraft. The objective function is to optimizing weight penalties on tardiness time. The model was solved to schedule the set of job to fixed due dates with WIP level was determined.

There are many discoveries around semiconductor wafer generation, such as Chang et al. (2003), Manufacturing Operations Control (PAC) technical analyst for an MTO/MTS mixed semiconductor fabrication organization. See Anavi-Isakow and Golany (2003) for a variation of the CONWIP standard for table projection.

3. Methodology

3.1. Conceptual Design Description

3.1.1. Problem definition

In this section, a picture of existing inventory issues related to TBS's production system was observed and analyzed. The weaknesses of production systems were discussed, as well as the current problems with which these types of systems are still struggling to achieve significant success.

3.1.2. Setting objective

During this period, the production value chain of TBS company will be analyzed. On that basis, the study will propose a more suitable production plan for the factory.

3.1.3. Model formulation

During this phase, a mathematical model is built based on the goal of minimizing total operating costs by rigorously determining the number of containers to be used and arranging production schedules on the appropriate backlog list. for different needs of customers. To test the validity of the proposed theoretical system, a simulation model with conceptual model, hypothetical input data is built and run to get the result.

3.1.4. Model translation

IBM CPLEX ILOG software will be applied to solve Mixed Integer Programming to design the production planning system by inputting math into the coding.

3.1.5. Verify the model

This process ensures that if the model behaves as intended and is accurate.

3.1.6. Data collecting

Based on the list of index and parameter, important data are collected.

3.1.7. Data analysis and processing

The next step is to modify the data, then the required data is verified and imported into the model to solve it. Collected data, along with building model, code is fed into CPLEX to find the optimal solution for the proposed production plan optimization.

3.1.8. Model validation

After a suitable model is built, the data is applied to check whether these results improve compared to the current situation of the company. If they match expectations, move on to the next step. Otherwise, a modification will be made to get the model and the final optimization result.

3.1.9. Result analysis

Input the collected data set into the final model to produce the results.

3.1.10. Conclusion and recommendation

According to the results from the previous step, there will be a suggestion for the business on how to do to improve the business. There should also be a discussion of how the proposed model can be upgraded.

3.2. Modeling

3.2.1. Assumptions

The method of establishing requirements for a backlog list when the number of containers is infinite is comparable to the flow shop sequencing problem that is limited to the phased plan. The explanation that the identical problem is limited to stage plans is that the CONWIP discipline expects the first-come, first-served discipline between machines on the generation chain. Because of this similarity, we realized that when all was done, tracking down an ideal arrangement was particularly difficult. Building construction requirements using a specific method is particularly appealing when managing long lead times. Note that if two consecutive things of redundancy deal with the same part, then a new set-up on a given machine may not be required between production of the parts. Therefore, when set-up is needed, they are usually inherited subordinates.

The paper makes some assumptions in proving the CONWIP generation framework. We accept that time is divided into time periods (days) and within these periods, into regular time and overtime. The need is known. Parts that are fabricated before the relevant demand is known is called inventory. This last assumption is reliable for "pull" reasons. There is a penalty for completing the generation before the request (FG inventory cost). Likewise, there is a penalty for overdue transfers. We accept that the raw materials are consistently approachable which makes sense since this is a temporary arrangement pattern. Preparation times are characterized and known as parts; sort time depends on the inheritance implemented which may exist. Holding costs are direct and known for all divisions (for both WIP and FG inventories).

We simply accept that the exchange cluster size is known in advance (remotely characterized for the model). Because they are consistent, the exchange lot size can be considered of the unit size without losing consensus.

3.2.2. Index

M	Number of machines.
I	Number of part types.
N	Number of containers.
T	Planning horizon [period].

3.2.3. Parameter

D_{it}	Demand for part i at the end of period t (boxes).
K	Number of entries in the backlog list; $\sum_{i=1}^{I}\sum_{t=1}^{T} D_{it}$.
h_i^w	Holding cost of WIP inventory for part i [$/period*box].
h_i^f	Holding cost of finish goods for part i [$/period*box].
b_i	Backorder cost for part i [$/period*box].
O	Overtime cost [$/hour].
r_{max_t}	Maximum available regular time in period t [hour].
o_{max_t}	Maximum available overtime in period t [hours].
p_{im}	Processing time of part i on machine m [mins].
τ_{ijm}	Sequence dependent set up time for switching from part i to part j on machine m [mins].
$BigM$	A large number.

3.2.4. Decision variable

x_{ik}	1 if the kth entry on the backlog list represents part type i. 0 otherwise.
z_{kt}	1 if the forecasted release time of the part associated with the kth entry on the backlog list falls in period t. 0 otherwise
f_{kt}	1 if the forecasted finish time of the part associated with the kth entry on the backlog list falls in period t. 0 otherwise
v_{ikt}	1 if the kth entry on the backlog list represents part type i and its forecasted release time falls in the period t. 0 otherwise
u_{ikt}	1 if the kth entry on the backlog list represents part type i and its forecasted finish time falls in period t 0 otherwise
y_{ijk}	1 if the kth entry on the backlog list represent part type j and a setup form part i to part j is required. 0 otherwise
S_{it}^{+}	Finish good inventory of part i at the end of period t (S_{i0}^{+} $given$) [boxes].
S_{it}^{-}	Backorders of part i at the end of period t (S_{i0}^{-} $given$) [boxes].

$\boldsymbol{W_{it}}$ WIP inventory of part i at the end of period t [boxes].

$\boldsymbol{r_t}$ Regular time used in period t [mins].

$\boldsymbol{o_t}$ Overtime used in period t [mins].

$\boldsymbol{s_{km}}$ Starting time of k representing part i on machine m [mins].

$\boldsymbol{f_{km}}$ Finishing time of k representing part i on machine m [mins].

3.2.5. Objective function

$$\text{Min } \sum_{t=1}^{T}\sum_{i=1}^{I}(h_i^f S_{it}^+ + b_{it}S_{it}^- + h_i^w W_{it}) + O\sum_{t=1}^{T} o_t \quad (1)$$

3.2.6. Constraint

$$S_{i,0}^+ - S_{i,0}^- + \sum_{k=1}^{K} u_{ikt} - D_{it} = S_{it}^+ - S_{it}^- \quad \forall i \in I, t = 1 \quad (2)$$

$$S_{i,t-1}^+ - S_{i,t-1}^- + \sum_{k=1}^{K} u_{ikt} - D_{it} = S_{it}^+ - S_{it}^- \quad \forall i, t > 1 \quad (3)$$

$$W_{i,0} + \sum_{k=1}^{K}(v_{ikt} - u_{ikt}) = W_{it} \quad \forall i \in I, t = 1 \quad (4)$$

$$W_{i,t-1} + \sum_{k=1}^{K}(v_{ikt} - u_{ikt}) = W_{it} \quad \forall i, t \quad (5)$$

$$\sum_{k=1}^{K} x_{ik} = \sum_{t=1}^{T} D_{it} \quad \forall i \in I \quad (6)$$

$$\sum_{i=1}^{I} x_{ik} = 1 \quad \forall k \in K \quad (7)$$

$$v_{ikt} = x_{ik}z_{kt} \quad \forall i, k, t \quad (8)$$

$$u_{ikt} = x_{ik}f_{kt} \quad \forall i, k, t \quad (9)$$

$$y_{ijk} = x_{i,k-1}x_{jk} \quad \forall i, j, k \quad (10)$$

$$x_{ik} + z_{kt} - 1.5 \le v_{ikt} \quad \forall i, k, t \quad (11)$$

$$x_{ik} + z_{kt} \ge 1.5 * v_{ikt} \quad \forall i, k, t \quad (12)$$

$$x_{ik} + f_{kt} - 1.5 \le u_{ikt} \quad \forall i, k, t \quad (13)$$

$$x_{ik} + f_{kt} \ge 1.5 * u_{ikt} \quad \forall i, k, t \quad (14)$$

$$x_{i,k-1} + x_{jk} - 1.5 \le y_{ijk} \quad \forall i, k > 1, t \quad (15)$$

$$x_{i,k-1} + x_{jk} \ge 1.5 * y_{ijk} \quad \forall i, k > 1, t \quad (16)$$

$$e_{km} = s_{km} + \sum_{i=1}^{I} x_{ik} * P_{im} + \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ijk} * \tau_{ijm} \quad \forall k, m, i \ne j \quad (17)$$

$$s_{km} = 0 \quad k = 1, m = 1 \quad (18)$$

$$s_{kM1}\begin{cases} e_{k-1,M1} + 1 \\ \max\left(e_{k-N\,;\,ML}, e_{k-1,M1}\right) + 1 \end{cases}$$
$$\forall k, \text{M1: initial machine, ML: last machine} \quad (19)$$

$$s_{km} \ge e_{(k-1)m} + 1 \quad \forall k > 1, m \quad (20)$$

$$s_{km} \ge e_{k(m-1)} + 1 \quad \forall k, m > 1 \quad (21)$$

$$\sum_{a=1}^{t} z_{k_1 a} \ge \sum_{a=1}^{t} z_{k_2 a} \quad \forall k_1 > k_2, \forall t \quad (22)$$

$$1 + \frac{\sum_{a=1}^{t}(r_a + o_a) - s_{km}}{\text{BigM}} \ge z_{kt} \quad \forall t, k, m \in \text{initial machine} \quad (23)$$

$$\frac{\left(\sum_{a=1}^{t}(r_a + o_a) - s_{km}\right)}{\text{BigM}} - \sum_{a=1}^{t-1} z_{ka} \le z_{kt} \quad \forall t, k, m \in \text{initial machine} \quad (24)$$

$$\sum_{t=1}^{T} z_{kt} = 1 \quad \forall t, k \quad (25)$$

$$\sum_{a=1}^{t} F_{k_1 a} \ge \sum_{a=1}^{t} F_{k_2 a} \quad \forall t, \quad (26)$$

$$1 + \frac{\sum_{a=1}^{t}(r_a + o_a) - e_{km}}{\text{BigM}} \ge f_{kt} \quad \forall t, k, m \in \text{last machine} \quad (27)$$

$$\frac{\left(\sum_{a=1}^{t}(r_a + o_a) - e_{km}\right)}{\text{BigM}} - \sum_{a=1}^{t-1} f_{ka} \le f_{kt} \quad \forall t, k, m \in \text{last machine} \quad (28)$$

$$\sum_{t=1}^{T} f_{kt} = 1 \quad \forall t, k \quad (29)$$

$$r_t \le r_{maxt} \quad \forall t \quad (30)$$

$$o_t \le o_{maxt} \quad \forall t \quad (31)$$

$$x_{ik}, y_{ijk}, z_{kt}, f_{kt}, v_{ikt}, u_{ikt} \in (0,1) \quad \forall i, j, k, t \quad (31)$$

$$S_{it}^+, S_{it}^-, W_{it}, r_t, o_t, s_{km}, f_{km} \ge 0 \quad \forall i, k, t, m \quad (32)$$

Objective function (1) is to minimize all existing costs including finished goods holding costs, backorder cost, WIP inventory holding cost, and overtime cost. We assume that the cost for regular time is independent of our decision. Constraints (2) - (5) enforce inventory balance for finish good and WIP inventories, respectively. Constraint (6) ensures that all demand will be satisfied during the planning horizon. (7) guarantees that one, and only one, part is assigned to each entry of the backlog list. Constraints (8) – (10) ensure the desired behavior of the binary variables v_{ikt}, u_{ikt}, and y_{ijk}, respectively. (11) – (16) a linear version of constraints (8) – (10). Constraints (17) the finishing time of entry k representing I on machine m is equal to its starting time on machine m plus its required processing time and set-up time on machine. Constraint (18) the first entry 1 in the part list starts its operation on the first machine at time 0. (19) Starting time of entry k other than 1 on the initial machine of the line depends on the availability of containers. The part is allowed to enter the production line only when the container is available. If container is available, the part will enter the line immediately after the previous entry finishes on the first machine. If there is no container available, even the initial machine is idle, the entry cannot enter the line. It must wait unit a container is empty entered the line prior to it. Constraints (20) – (21) Entry k > 1 representing part i can only start on machine m>1 until finishes its proceeding operations on the previous machine, m-1, and the entry prior to it finishes the operations on machine m. Constraints (22) – (29) transform the results for release and finish times to the binary variable z_{kt} and f_{kt}. Constraints (30) and (31) introduce upper bounds on the regular time and overtime, for each time period. The reason r_t may actually be less than r_{maxt} is that there is a penalty for completing work early (finished good costs).

3.3. Data collection

Data collected from Thai Binh Shoe Company's database is subject to their permission and privacy policy, including:

- Customer demand,
- Processing time for each type of machine,
- Storage costs and penalties for overdue delivery,
- Overtime cost,
- Et cetera.

Figures are provided by estimates based on customer order lists and production planning for three business days of 3 types of sports shoes with codes SK-44369, SK-3BG72, SK-W1426, respectively.

The shoe production process of TBS Company is very straightforward with 5 machines in the system having its own limited capacity. The material begins to be processed there during cutting to produce amorphous WIP. These WIPs are then synchronized into the design shape. Next, the logo or pattern is printed before sewing to create the shape. Finally, the upper is assembled with the outsole to form a complete shoe.

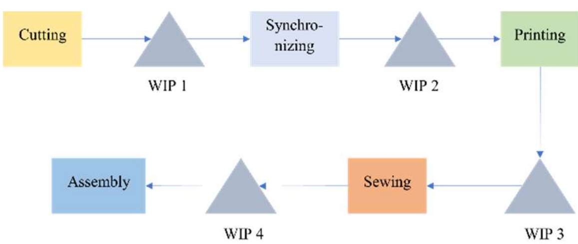


Figure 1. Shoe production process in TBS Company

4. Result analysis

4.1. Mathematical model result

Based on real data collected about production needs and operating productivity of the factory, the thesis solved the problem of Optimizing production costs with the CONWIP system of IBM ILOG CPLEX Optimization Studio 12.10. This leads to the optimal solution for backlog production schedule design, minimizing all associated costs (storage costs, backorder costs, overtime costs) This section describes how to Strict container quantity and production schedule to achieve optimal cost

4.1.1. Number of containers or WIP level

Table 1. Result of total cost and makespan with different WIP level

WIP level	Total cost (VND)	Makespan (Minutes)
W = 1	-	-
W = 2	-	-
W = 3	30,563,330	1795
W = 4	19,603,330	1789
W = 5	19,603,330	1789
W = 6	19,516,670	1787

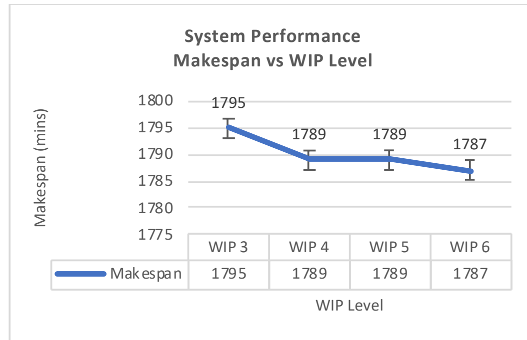


Figure 2. Makespan vs WIP level in production line

The makespan decreases as the WIP level increase. As the number of containers or WIP level increases, the waiting time between machines and products entering the system decreases. However, once the WIP reaches certain point, adding more containers or increasing WIP levels do not improve the system makespan. The line chart above illustrates at N=3, makespan is quite high, while N=4 to 6, makespan is significantly reduced to maintain stability, and the difference is negligible.

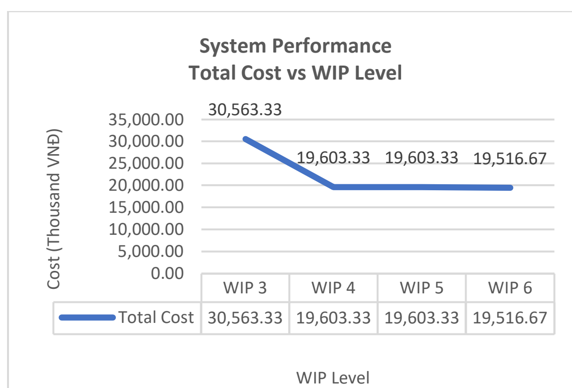


Figure 3. Total cost vs WIP level in production line

Similar to makespan, at N=3, the total cost is considerably high and by 1.56 times than at N = 4. Whereas the WIP reaches point from 4 to 6, the total cost is almost remained stable. So, in this TBS production line, the best WIP level or number of containers is 4.

4.1.2. Quantity of inventory and backorder

As the mathematical model mentioned, the total cost of the objective function is considered by 4 factors. In addition to total working time, three other factors were also considered: lost sales, inventory cost of finished goods and WIP.

Number of WIP inventory at each WIP level:

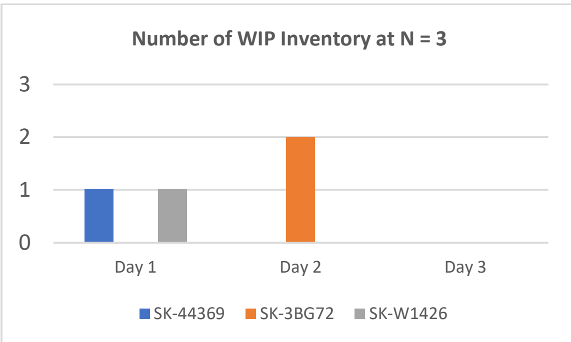


Figure 4. Number of WIP inventory at N = 3

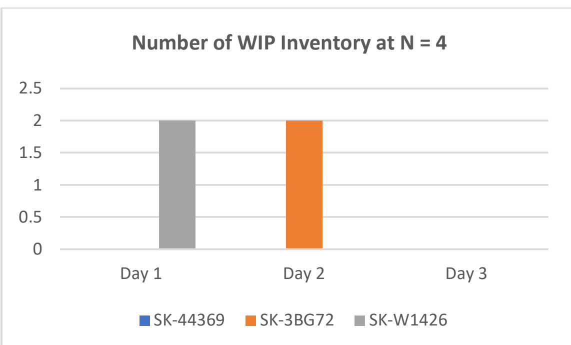


Figure 5. Number of WIP inventory at N = 4

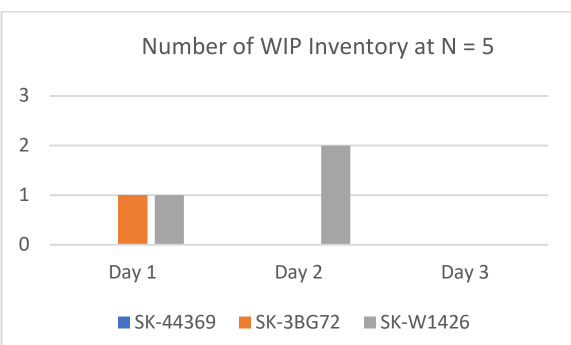


Figure 6. Number of WIP inventory at N = 5

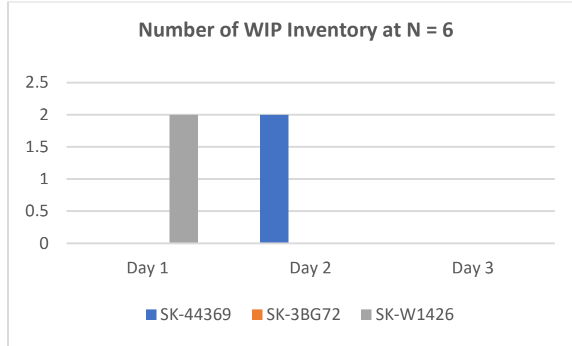


Figure 6. Number of WIP inventory at N = 6

The results of WIP inventory at four different WIP levels show that the total inventory of four levels is the same. However, the amount of WIP inventory for each product type varies. This depends on the backlog list when it goes into production.

The bar chart below depicts a significant difference between the two WIP levels of finished goods inventory and backorders within three working days. At N = 4, most products are produced just enough to order in each period, whereas at N = 3, lead time and production scheduling causes excess inventory in some products and others are late for delivery.

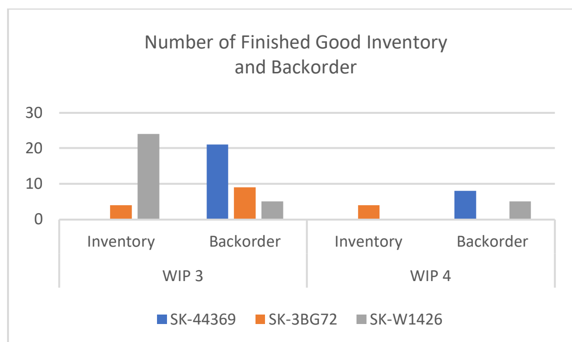


Figure 7. Number of finished goods inventory and inventory

4.1.3. Cost from output model

By considering the level of WIP for flexibility in manufacturing operations, the results of each cost are shown below:

Table 2. Comparison type of cost in two WIP level

Cost	WIP 3	WIP 4
Holding cost of finish good	6,540,000 VND	1,020,000 VND
Holding cost of WIP	460,000 VNĐ	460,000 VNĐ
Backorder cost	8,180,000 VND	3,000,000 VND

Overtime cost	23,075,000 VND	22,685,000 VND
TOTAL	**38,255,000 VND**	**27,165,000 VND**
SAVING 11,090,000 VND		

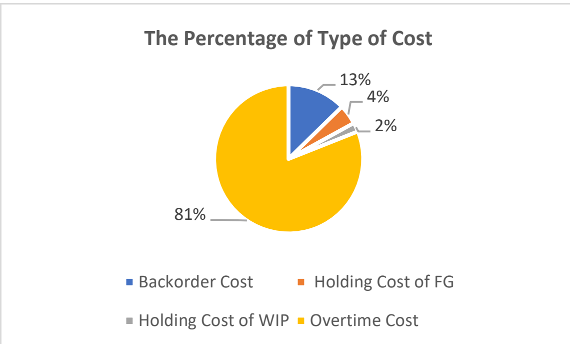


Figure 8. The percentage of type of cost

The chart depicts the percentage of each type of expense. Overtime costs account for the highest percentage of 81% compared to other types of costs, this depends on the number of workers and total production time. However, this cost is almost fixed and difficult to change. To reduce production time, the factory must consider increasing throughput, and it can affect a number of factors such as product quality. In addition, manufacturers can consider increasing the number of containers, but the WIP level only reaches a certain level. This is the optimal level of makespan as well as the order in which the backlog goes into production.

5. Conclusion and Recommendation

5.1. Conclusion

Utilizing and sharing resources to fulfill orders is necessary for large-scale manufacturing companies like TBS today. To address the factors listed in the previous section, this thesis has developed a mathematical programming model to analyze the CONWIP system in this study. The model was developed for a serial CONWIP production line. The solution of that model defines the sequence of jobs to be processed in CONWIP line, the cost types, and the job processing time. The model considered backorder, holding, and overtime costs in balance with the workload and time required to complete the horizontal plan. In addition, the number of containers can also be determined when solving the model, the goal is to minimize the gap in production time, meet demand in time and at the same time achieve the minimum WIP. Calculation results have shown a clear relationship between WIP levels.

The efficiency of controlling the number of containers and scheduling orders on the backlog greatly reduces costs. This will help reduce labor costs and other costs related to warehousing, thereby increasing efficiency, and increasing profits for the company.

5.2. Limitation

The current model is only applicable to the product demand of the production line on a set of factories, because when solving such mathematical programming models for large-scale problems, the search algorithm limited generalization and branch based cannot give optimal or near-optimal solutions in acceptable computation time. The time to solve a problem is quite long, so it is difficult to apply to long term production options.

The use of CPLEX in finding the optimal solution of a mathematical model also has some limitations. Mathematical models can be used to solve in an exact manner with a globally optimal solution; however, CPLEX and mathematical modeling are not suitable for implementing a large data set. Algorithms and heuristics are required to improve speed and meet the requirements of solving large problems.

5.3. Recommendation

For future research in terms of modeling, there are several interesting areas where existing models can be extended:

In this study, the processing time and the setup time were assumed to be known and determined. This could be extended in the future to have a defined processing time with random machine failures and repairs.

In this study, the WIP level was set at the limit for a production line or the entire production system. It would be interesting to move the WIP levels down for each part type. This can be modeled by having CONWIP constraints for each part.

References

Anavi-Isakow, S. and Golany, B. (2003). Managing multi-project environments through constant work-in-process. *International Journal of Project Management*, 21, 9–18

Bard and Golany (1991). Determining the number of kanbans in a multiproduct, multistage production system. *The International Journal of Production Research,* 29(5), 881-895

Cao, D. and Chen, M. (2005). A mixed integer programming model for a two-line CONWIP-based production and assembly system. *International Journal of Production Economic,* 95, 317–326.

Chang, S.H. et al. (2003). Heuristic PAC model for hybrid MTO and MTS production environment. *International Journal of Production Economics*, 85, 347–358.

Fowler, J. W., Hogg, G. L. and Mason, S. J. (2002). Workload control in the semiconductor industry. *Production Planning & Control: The Management of Operations,* 13(7), 568-578

Gaury, E. G. A., Pierreval, and Kleijnen, J. P. C. (2000). An evolutionary approach to select a pull system among Kanban, Conwip and Hybrid. *Journal of Intelligent Manufacturing,* 11, 157–167

Giovanni, R. (2015). Design and Simulation of CONWIP in the Complex Flexible Job Shop of a Make-to-Order Manufacturing Firm*. International Journal of Industrial Engineering Computation,* 6 (1), 117-134.

Herer, Y. T. and Masin, M. (1997). Mathematical programming formulation of CONWIP based production lines; and relationships to MRP, *International Journal of Production Research,* 35(4), 1067-1076

Hopp, W. J. and Roof, M.L. (1998) Setting WIP levels with statistical throughput control (STC) in CONWIP production lines, *International Journal of Production Research,* 36:4, 867-882

Luh, P.B., Zhou, X. and Tomastik, R.N. (2000). An Effective Method to Reduce Inventory in Job Shops. *IEEE Transactions on Robotics and Automation,* 16(4).

Rees, L. P., Philipoom, P. R., Taylor III, B. W. and Huang, P., Y. (1987). Dynamically Adjusting the Number of Kanbans in a Just-in-Time Production System Using Estimated Values of Leadtime. *IIE Transactions,* 19(2), 199-207

Remco, G. and Riezebos, J. (2010). Workload Balancing Capability of Pull Systems in MTO Production. *International Journal of Production Research*, 48(8), 2345–2360.

Spearman, M., L. and Zazanis, M., A. (1992) Push and Pull Production Systems: Issues and Comparisons. *Operations Research* 40(3), 521-532

Spearman, M.L., Woodruff, D.L., Hopp, W.J. (1990). CONWIP: a pull alternative to Kanban. *International Journal of Production Research*, 28 (5), 879–894

Stefan, H. Schimmelpfeng, K. and Stolletz, R. (2011*)*. Setting Inventory Levels of CONWIP Flow Lines via Linear Programming*. Business Research Journal*, 4 (1), 98–115

ENHANCEMENT OF SALES FORECAST PROCESS FOR A JEWELRY RETAILER USING SARIMA AND WINTERS TRIPLE EXPONENTIAL SMOOTHING

Thuy Duong Nguyen[1], Ngoc Hien Do[2]

[1] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam,* ntduong.sdh19@hcmut.edu.vn

[2] *Department of Industrial Systems Engineering, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam,* hienise97@hcmut.edu.vn

Abstract. This paper would present an enhancement of the forecast process that has been used in a jewelry retailer. Comparing SARIMA and Holt-Exponential Winter's Smoothing techniques in order to provide high-accuracy customer transaction forecasts, which would help the retailer to prepare sufficient resources capacity. Due to basic characteristics of historical time series, appropriate theoretical forecast models are used. They would be ranked by comparing forecast accuracy and forecast bias to find out which one is the best forecast model for the case study. After applying the solution, the forecast accuracy was increased by 10%. The results of this study could be applied to other jewelry retailers, or retailers in other industries, with some necessary modifications.

Keywords: Sales Forecast; Forecast Process; Forecast Accuracy; Jewelry Retailer

1. Introduction

Inventory management is an important aspect of running a fashion store. Proper retail inventory management, which aids in the balance of supply and demand, is heavily depended on accurate demand forecasting. Sales forecasting is the process of predicting future sales using historical data. It is becoming increasingly important in a commercial enterprise's decision support system as a result of competitiveness and globalization (Tiaojun and Qi, 2008). Effective sales forecasting may assist decision-makers in calculating production and material costs as well as determining the selling price (Zhan-li et al. 2008) leading to reduce over-stocking and under-stocking costs. It is an evidence from the preceding discussions that fashion retail sales forecasting is a critical topic in practice.

Several techniques such as Winters exponential smoothing, Box Jenkins ARIMA model, and multiple regression have been devised and widely utilized to account for these patterns during the previous few decades. In sales forecasting, statistical time series analysis methods like ARIMA and SARIMA are commonly used (Box, et al. 2015). These methods are simple and easy to implement. The results may be computed fast since they use a closed form expression for forecasting. Meanwhile, Holt Winter's Exponential Smoothing is accurate in forecasting seasonal time series data, regardless of whether the pattern displays a trend (Kalekar, 2004).

This study is to evaluate the forecasting performance of Winters models and ARIMA models, which are applied to forecast retail sales of 1 distinct category of jewelry company as a case study. In addition, multiple cross-validation samples would be evaluated via the RMSE (root mean squared error), MAPE (Mean absolute percent error), FA (forecast accuracy), and FB (forecast bias) to conclude which forecast method is the best for this circumstance (Wei, 2006).

2. Literature review

The SARIMA technique, or Seasonal Autoregressive Integrated Moving Average, is a Time Series forecasting method for stochastic model data with seasonal data patterns. George Box and Gwilym Jenkins publiced the ARIMA model in the 1970s, with applications in time series analysis and forecasting. Box and Jenkins (1970) characterized underlying theories, then Reinsell (1994) is sophisticated to simply understand and apply. There are four steps of SARIMA modeling (Bouzerdoum et al. 2013). (1) The model identification phase identifies variables to validate the time-series stationarity; it also finds the best auto-regression and moving average combination. (2) The model estimation phase evaluates the models found in the previous step and selects the most efficient one; (3) The model validation phase evaluates the chosen model's precision; (4) The model forecasting phase predicts the series future data and provides a confidence interval. Winters Smoothing technique is a forecasting method that uses an exponential smoothing strategy based on previous forecasting results, especially seasonal data patterns. Holt-Exponential Winter's Smoothing technique has two primary models: multiplicative and additive. This model's selection was based on the seasonal pattern (Yang et al. 2017).

2.1. SARIMA model

SARIMAX (p,d,q) (P, D, Q)S (X), where X is the vector of external variables, is a SARIMA model with external variables.

With:

- p, d, q : The non seasonal part of the model
- (P, D, Q)(S): The seasonal part of the model
- S : Number of periods per season

The general SARIMA equation is shown as below:

$$\Phi_p B^S \phi_p(B)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_q(B^S)a_t \quad (1)$$

2.2. Winters model

Various scenarios were applied by Winters' exponential smoothing model, in particular data with trend and seasonal. Chen and Winters (1966) used it in the context of predicting peak demand for an electric utility company. The Winters model was demonstrated to be a reliable model effective for forecasting individual product sales, business sales, income statement items, and aggregate retail sales (Alon et al. 2001). Multiplicative Seasonal Method (Multiplicative Seasonal Method) for seasonal variation of data that has grown/reduced (fluctuation) and Additive Seasonal Methods for stable seasonal variation which are two divisions of this method.

2.3. Measures of Accuracy

According to all accounts, several metrics are applied to various research to assess the predicting performance of models (Wang et al. 2018). The root mean square (RMSE), mean absolute percent error (MAPE) are usually selected to appraise the forecasting performance of prediction models, and their expressions are shown as follows:

$$MAPE = \frac{\sum_{t=1}^{n}\frac{|\varepsilon|}{Y_t}}{n} = \frac{\sum_{t=1}^{n}\frac{|Y_t-\hat{Y}_t|}{Y_t}}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\varepsilon^2}{n}} = \sqrt{\frac{\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2}{n}} \quad (3)$$

where y (t) and Y (t) are the predicted values, observed values, respectively, and n is the size of the testing set. Smaller RMSE and MAPE values indicate more precise prediction outcomes.

Using the Python language, the best fit model was chosen based on the lowest value of the Akaike Information Criterion (AIC) and the Schwartz Bayesian Information Criterion (BIC):

$$AIC = ln\hat{\sigma}^2 + \frac{2}{n}r \quad (4)$$

$$BIC = ln\hat{\sigma}^2 + \frac{\ln(n)}{n}r \quad (5)$$

Besides, instead of looking at the error of the forecast, the accuracy of the forecast (FA), the forecast bias (FB) was consodered as below equation:

$$FA = 1 - MAPE = 1 - \frac{|Actual - Forecast|}{Actual} \quad (6)$$

$$FB = \frac{Actual - Forecast}{Actual} \quad (7)$$

With:

FB >0: lost sale, low level of response of the goods

FB <0: high inventory level

Where p^B (AR Non Seasonal), Φ_p B^S (seasonal AR polynomial), $(1-B)^d$ (differencing non seasonal), $(1-B^S)^D$ (differencing seasonal), θ_q (B) (MA non seasonal), Θ_q (B^S) (MA seasonal)

3. Methodological Procedures

Because the "researcher has no influence on the independent variable, which is the assumed factor of the event, because it has already occurred," this is an ex-post facto study (De Gooijer et al. 2006). It is a sectional study in terms of temporal cutting. The nature of the objective of study as well as the data procedure supported a quantitative approach to the research problem. The product category, which consists of 20 SKUS (Stock KeepingUnits) as well as accounts for 24% of overall sales in the local market, the leading area of sales number in nation, was studied.

Individual reviews of the various goods reveal that they have no suitability in management decision-making. As a result, in a particular time series, numerous of comparable goods can be aggregated and evaluated together based on purposeful criteria (Zotteri et al. 2005).

Two linear demand of quantitative forecasting models would be examined for a full comparison of distinguished methodologies: the model of Winters (W) and SARIMA model.

The methodology used in this research includes three phases as follows.

Phase 1: Defining the improvement area

In this phase, by comparing collected data to key performance indicators (KPIs) of the retailer, gaps in the forecast accuracy, the forecast bias of the in-use forecast procedure are pointed out clearly so as to establish the need for improvement in this paper. After that, the improvement area in this paper is clarified.

Phase 2: Proposing the solution

Based on the results of defining the improvement area, a solution that is considered adapting the forecast requirements of the retailer in terms of the accuracy is proposed.

Phase 3: Analyzing experimental results

Eventually, experimental results of simulating the forecast procedure on a historical data set are used to demonstrate the usefulness and efficiency of the solution.

Defining the improvement area

The jewelry retailers studied in this paper is a jewelry joint stock company. It is well-known for expensive product groups such as gold, silver, gems, business gifts, fashion accessories, souvenirs, watches, buying and selling gold bars.

Based on the collected data of average accuracy and bias of sales forecast of 12 months in 2018 (Figure 2), all forecast values at that time were not satisfied KPIs of the forecast department that the accuracy must be at least 80%, and the bias cannot exceed 20%.

Therefore, the research would focus on creating an enhanced process to solve inappropriate in-use models aiming at improving forecast accuracy.

4. Proposing the solution

4.1 Data analysis

Firstly, the monthly sales data (from January 2012 to November 2018) were collected to find out the trend and check for stationarity. Group is TSCZ in Ho Chi Minh area which is the highest proportion. Thus, TSCZ is chosen in this study. The data collected is shown in Table 1 as following.

Table 16. Sales data

Time	Group	Region	Quantity (pcs)
2012 - 01	TSCZ	HCM	1418
2012 - 02	TSCZ	HCM	2264
2012 - 03	TSCZ	HCM	1923
		……	
2018 -11	TSCZ	HCM	7915

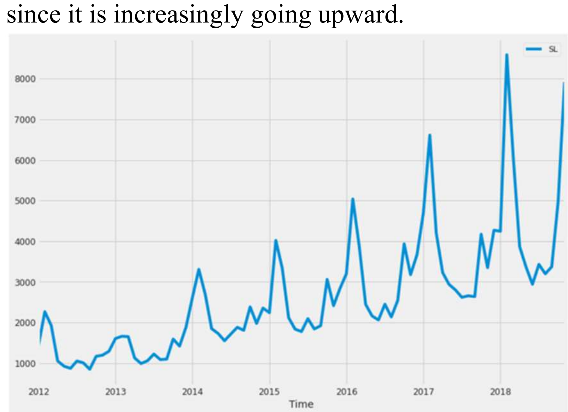
4.2 Model selection

Two approaches would be used in the study which are SARIMA model known as Box – Jenkins and and the Holt-Winters Triple Exponential smoothing.

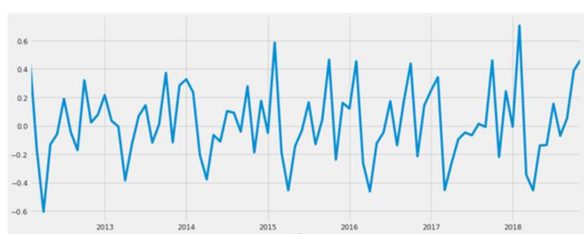
4.2.1 Box-Jenkins/SARIMA Model

4.2.1.1 Model Indentification

Fig. 1 (a) shows that the data (as in Table 1) has a trend since it is increasingly going upward.



a) Plot of monthly sales



b) Plot of time series sales

Figure 6. Non-stationary and Stationary plots

Data show trend and seasonality (as on Fig. 1 (b)), so it also needs to be defined the seasonal index. Data were converted by using log and first difference as well as re-evaluated the trend. Augmented Dikey Fuller (ADF) test to check stationarity, and the results indicated stationarity was used as shown in the Table 2.

Hypothesis test for the Augmented Dickey-Fuller t-test is as followings:

H_0: The data is not stationary which needs to be differenced

H_1: The data is stationary which doesn't need to be differenced

The test results are summaried as in Table 2. The p-value is less than 0.05 thus H_0 was rejected. It means that data are clearly stationary.

Table 17: Dickey – Fuller Test Result

Results of Dickey-Fuller Test	
Test statistics	-3.420844
p-value	0.010268
#Lags used	11.00000

4.2.1.2 Model Estimation and Selection

ACF and PACF were shown in Fig. 2. After 12 lag (13, 25), seasonal spikes at ACF and PACF are visible. It indicates that the series do take seasonal differences into account. The non-seasonal part of the PACF shows that there is a spike under the dashed line at lag 2 and no spike till lag 5, 13 and then a discontinuation. The AR (2) were suggested with already variaty in data, so the model (p, d, q) would be (2, 1, 2). Therefore, seasonal difference does not required. the appropriate SARIMA model for forecasting sales and hence inflation would be (2,1,2)(P,D,Q)(12).

The ACF and PACF plots as followings would indicate probable AR and MA.

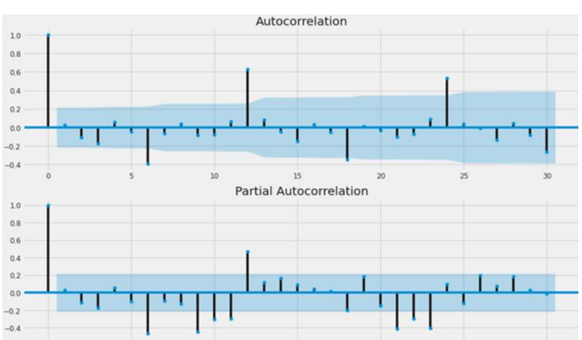


Figure 7. Plot of differenced transform monthly sales

After identifying and estimating (p,d,q)(S) by ACF, PACF as above, we use Python language to re-examine each index (p,d,q) and find out (P,D,Q) which is applied and calculated AIC. The best fit model would be determined by doing a stepwise search to minimize the AIC.

Table 18: ARIMA index (p,d,q) testing result

Performing stepwise search to minimize aic	
ARIMA(0,1,0)(0,1,0)[12]	: AIC=-6.150, Time=0.03 sec
ARIMA(2,1,2)(2,1,0)[12]	: AIC=-48.609, Time=2.31 sec
ARIMA(0,1,2)(2,1,0)[12]	: AIC=-50.917, Time=0.96 sec
Best model: ARIMA(0,1,2)(2,1,0)[12]	
Total fit time: 129.267 seconds	

As shown in the Table 3, the model SARIMA (0,1,2)(2,1,0)(12) has the least AIC (-50.917) and times (0.96s) values. Therefore, SARIMA ((2,1,2)(2,1,0)(12) was not a selected model due to ACF, PACF plot which has higher AIC (-48.609) and times (2.31s) in comparison.

4.2.1.3 Forecasting

Forecasted values for demand in January, February are 4477 and 7763 respectively.

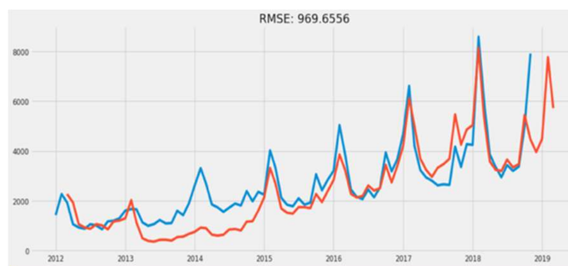


Figure 8. Plot of actual and forecasted values with SARIMA method

4.2.2 Triple Exponential Smoothing (Winters Method)

4.2.2.1 Model Estimation and Selection

The steps of forecasting with Winter's Exponential Smoothing method include analyzing the data to see if it contains trend and seasonal elements by looking at the pattern formed.

The three components of the smoothing triplet (E, T, S) were defined by Hyndman (Hyndman, 2008) as error, trend, and seasonality. In this paper, ETS function of Python for automated model selection was used. The accuracy was a criteria for choosing the best model.

The selected model is suggested by ETS, as shown on Figure 4, which indicates that the model contains additive errors, additive trend, and seasonality. They are indicators of Winters linear with additive errors. These confirms are the same with confirming from SARIMA model results.

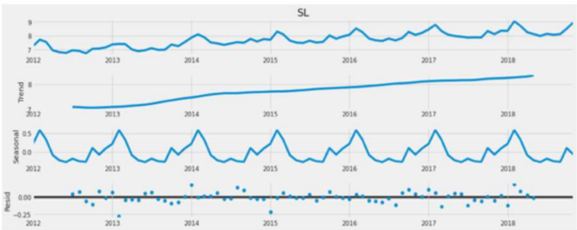


Figure 9. Decomposition by ETS

4.2.2.2 Forecasting

Forecasted values for demand in January, February 2019 are 6573 and 5277 respectively.

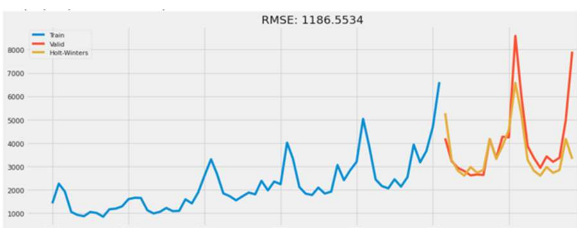


Figure 10: Plot of actual and forecasted values with Winters method

5. Discussion

Forecast sales transactions at a retail company were forecasted using SARIMA and Holt-Exponential Winter's Smoothing. According to the data properties, the index of SARIMA will be modified which are compared with Winter to evaluate the best fit model month on month and also be a new approach for staff at company.

Table 19: January and February Forecast comparison

Model	January Forecast	February Forecast
SARIMA	4477	7763
Winters	6573	5277

The forecast model accuracies were calculated, tested, and compared by means of MAPE and RMSE factors. From Table 5 below, the value for SARIMA model (0,1,2)(2,1,0)(12) were 0.03 and 0.21 for MAPE, RMSE respectively, while the Winters Triple Exponential smoothing model gave value of 0.21 and 1186.60 as value for MAPE and RMSE. Based on the accuracy parameters selected, this study indicated that the SARIMA models outperform more than the Winters method.

The research results are also get the agremence with Udom. P and Phumchusri. N (2014) who confirmed that ARIMA (1,0,1)(1,0,1)(12) was a better model in comparison with Holt-winters model based on mean Absolute Percentage Error factor. Nevertheless, conclusion from Puthran, et al. (2014) found that Holt Winters was less precise than SARIMA Model as shown in Table 5.

Table 5: SARIMA and Holt-Winters models comparision

Model	MAPE	RMSE
SARIMA	**0.03**	**969.70**
Winters	**0.21**	**1186.60**

Finally, Analysis of Variance FA, FB to evaluate the actual sale numbers for January and February 2019 which against the predicted values by using SARIMA and Winters in order to see the difference was done. The following results are shown in Table 6.

Table 20. FA, FB comparison between SARIMA and Winter method

Time	Quan-tity	Winters	Sarima	FA_W	FA_S	FB_W	FB_S
1/1/19	5734	4536	5542	79.1 %	96.6 %	- 20.8 %	- 3.4 %
2/1/19	9471	6625	11316	69.9 %	80.5 %	- 30.1 %	19.5 %

Both FA and FB of SARIMA forecast method have the better result:

- Forecast accurancy of Sarima is higher 10% than Winters
- Forecast bias of Sarima is lower 10% than Winters

6. Conclusion

Sales forecasting is becoming increasingly important in a commercial enterprise's decision support system as a result of competitive market and globalization.

The development of more accurate and timely sales forecasting systems has emerged as a prominent research subject.

This paper has considered to find out the suitable forecast sales model for a case study with using SARIMA and Holt-winters (Triple exponential Smoothing). The study results indicates that both models give almost similar results. However, when FA, and FB was used as criteria to compare, SARIMA model seemed more accurate in terms of FA, and FB values. The findings would assist in more accurate financial planning and budgeting when the demand forecast was done better.

The case study was shown that forecasted sales for January 2019 is 4467, February 2019 is 7764 while the actual inflation rate for January and February 2019 are 4020 and 6896. The suggested solution improved forecast accuracy around 10%. Actually, the data attributes should be analyzed before the optimal prediction technique is used. Planners can establish the best approach for deciding the time and selection of which products need to be exhibited or placed first in the warehouse based on precise forecasting findings or estimations. The study could be further enhanced by comparing with other forecasting techniques such as artificial neural networks in order to obtain better accuracy.

References

Alon, I., Qi, M. and Robert, J. S. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of retailing and consumer services,* 8(3), 147-156.

Bouzerdoum, M., Mellit, A. and Massi Pavan, A. (2013). A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy* 98, 226-235.

Box, G. and Jenkins, G. (1970). Time Series Analysis-Forecasting and Control. *San Francisco: Holden Day,* 553.

Box, G. E. P., et al.(2015). Time series analysis: forecasting and control. *John Wiley & Sons.*

De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting." *International journal of forecasting,* 22(3), 443-473.".

Hyndman, R. et al. (2008). Forecasting with exponential smoothing: the state space approach. *Springer Science & Business Media.*

Jenkins, G. M. (1994). Time series analysis. *Prentice Hall*.

Kalekar, P. S. (2004). Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi school of information Technology*, 4329008(13),1-13.

Tiaojun, X and Qi, X. (2008). Price competition, cost and demand disruptions and coordination of a supply chain with one manufacturer and two competing retailers. *Omega*, 36(5), 741-753.".

Wang, J. et al. (2018). A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. *Energy Conversion and Management*, 163, 134-150.

Wei, W. W. S. (2006) Time series analysis. *The Oxford Handbook of Quantitative Methods in Psychology,* 2.

Yang, Y., Yu, H. and Sun. Zh. (2017). Aircraft failure rate forecasting method based on Holt-Winters seasonal model. *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*.

Zhan-Li, S. et al. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1), 411-419.

Zotteri, G., Matteo, K., and Federico C. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics,* 93, 479-491.

HETEROGENEOUS FLEET VEHICLE ROUTING PROBLEM WITH TIME WINDOW: A CASE STUDY IN DKSH VIETNAM

Nguyen Van Chung[1], Ngo Thanh Phuong Anh[1] and Luong-Nhat Nhieu[2*]

[1]International University – Vietnam National University HCMC, Hochiminh city 70000,

nvchung@hcmiu.edu.vn (N.V.C.)

anhngotp@gmail.com (N.T.P.A)

[2]National Kaohsiung University of Science and Technology, Kaohsiung 80778,*

nnluong.iem@gmail.com (L.-N.N.)

Abstract: In this article, a company's healthcare business unit encountered a traffic problem that should be resolved by creating vehicle routes. According to the characteristics of the company, the problem is classified as a vehicle routing problem of a heterogeneous fleet with a time window, and the capacity and the number of vehicles is limited. Then a mathematical method is proposed and IBM CPLEX OPTIMIZER is used to solve the constraints. In order to evaluate the performance of the proposed method, experiments were conducted using real company data, and then compared with actual costs. It has been observed that the proposed method can be used to obtain reasonably better results for this problem. In addition, some ideas were put forward to actually implement the model and its effects in practice.

Keywords: Vehicle routing problem, Heterogeneous fleet of vehicle, Time window, Symmetric distance matrix.

1. Introduction

Nowadays, the efficiency of logistics activities is recognized as one of the factors that have the most profound influence on the overall performance of the business. The management of this logistics operation mainly includes managing and improving the process of inventory, import and export and transportation. Transportation problems with large volumes, high product diversity, limited vehicles, complex transportation networks, and time window constraints have always been a challenge for logistics managers. In search of a suitable answer to this challenge, more and more different approaches and variations of the vehicle routing problem (VRP) have been introduced by researchers in recent decades (Elshaer & Awad, 2020; Pan, Zhang, & Lim, 2021; Vidal, Laporte, & Matl, 2020). Among them can be mentioned mathematical optimization models that are developed increasingly complex with more and more real-world factors integrated. To continue this journey, as the main contribution of this article, a variant of the mathematical optimization model for heterogeneous fleet VRP was developed. Moreover, this proposed model is then applied to the actual operation of logistics enterprises in Vietnam as a validation process for later model improvements.

This paper is organized as follows: The next chapter introduces previous work regarding this topic. Chapter 3 presents methodology and chapter 4 is the results comparing to actual data, implementation and its impact. Finally, the last chapter offers conclusions and suggestions for future research.

2. Literature review

Hsu's research in 2007 aimed to determine the optimal transport routes within the constraints of time and number of vehicles for the perishable food distribution problem. This study uses stochastic mathematical models to solve vehicle routing problems with time window (VRP-TW). The results of the study show a significant share of inventory and energy costs in the system's cost structure. In addition, the authors also show the trade-off between inventory costs and fixed costs through adjustments for the number of vehicles. With stochastic time travel considerations, the results imply that time constraints become more difficult to satisfy than themselves under deterministic conditions (Hsu, Hung, & Li, 2007). In another study, published by Hande Öztop et al., mixed-integer programming and constraint programming models were introduced to identify optimal and near-optimal solutions for VRP-TW. In this study, heterogeneous vehicles are used to distribute valuable products such as cash to Automated Teller Machines. The resulting models provide safe routes with minimal travel time for heterogeneous vehicles. The limitation of the study mentioned by the authors is the small number of clients as a typical limitation of exact solution methods (Öztop, Kizilay, & Çil, 2020). In the global trend of reducing emissions, electric vehicles are also considered as a new mode of transport thereby creating another variant of VRP, electric VRP (EVRP). A study that developed an optimization model for EVRP with fuzzy uncertainty parameters was performed by Zhang et al. The above fuzzy uncertainty parameters describe the uncertainty conditions in EVRP such as energy consumption, travel time, service time, etc. In addition, the authors also propose an integrated algorithm with the core ALNS algorithm to enhance the solutions as well as optimize CPU time (Zhang, Chen, Zhang, & Zhuang, 2020). In research published in 2020, Hongqi Li considered the VRP-TW problem with modern unmanned aerial vehicles (UAVs). This study aims to optimize the delivery route of UAV-van vehicle complexes. The solution to the above problem was determined by the authors by the adaptive large neighborhood search algorithm. The results of this study are a reference for businesses about the feasibility of a new transportation mode (Li, Wang, Chen, & Bai, 2020). In 2020, Daniel proposed two different models for VRP-TW with limited time multiple trips. The numerical experiments of the study show that the proposed model is more efficient than the existing models mentioned by authors (Neira, Aguayo, De la Fuente, & Klapp, 2020).

Distribution networks are expected to become significantly different as distributors adopt customer picking through a network of locker facilities. A linear programming paradigm has been developed by Redi et al. to find a solution to VRP with such locker facilities. The numerical results of this study also demonstrate the advantages of customer picking strategy over home delivery strategy through the reduction of both important criteria, total transportation cost and number of vehicles (Redi et al., 2020). In Chen's research in 2020, a new optimization algorithm was developed for the problem of food distribution under conditions influenced by Covid-19, which was developed on the basis of the integration of an artificial bee colony algorithm with tabu search. The performance of this proposed algorithm shows that artificial bee colony algorithms can be improved through the tabu search mechanisms. The numerical results also show that the number of vehicles will increase if distribution network operators want to increase the delivery speed. In addition, the maximum working time is extended if a higher delivery rate is required (Chen, Pan, Chen, & Liu, 2020). In 2021, a study evaluating VRP solutions that trade-off between the three pillars of sustainable development was conducted by Hassana Abdullahi et al. The solutions in this study are determined through a metaheuristics algorithm that reconciles the advantages of a greedy local search method and a biased randomized savings heuristic. The three sustainability pillars are integrated in this study in the form of objective functions, which are costs representing economic factors, accident risks representing social factors, and CO2 emissions representing environment factors (Abdullahi, Reyes-Rubiano, Ouelhadj, Faulin & Juan, 2021).

3. Methodology

First, the problem and its objective are defined. Then, research papers related to the problem are searched and review. Next, based on research papers and reality, a model to solve the problem is construct and data is collected and analyzed. After that, by using the analyzed data as input, the model will be run and created a result. The result is then validated by comparing with the reality. If the outcome is not better than reality, the model will go back to problem modelling, data analyzing and running model and result validating again until it is better than reality. Finally, the conclusion and suggestion are provided based on the better outcome.

This paper proposes a mathematical model to solve a real-life heterogeneous fleet vehicle routing problem with time windows. The problem consists in determining, each day, how to allocate the trucks from depots to all customers, the amount to be delivered in each truck to each customer, which one is the best route and the suitable time for attending each node, with the aim of minimizing the total transportation cost (summation of fixed and operating costs), satisfying the customers demand and respecting all the problem constraints (vehicle number and feature, and time windows). The difference of this model comparing to the papers is, in this paper, a linearization method is introduced to transform two non-linear constraints into linear constraints. It also considered different types of products with different space, which both papers did not take into account. A directed graph G = (N, A) is given, where $N = \{1, 2, \dots, n\}$ is the set of n nodes. Node 1 represents the depot while the remaining node set $N\backslash\{1\}$ corresponds to the customers from node 2 to node n. Node set N has 3 indexes: i, j, $k \in N$. Each pair of location (i,j), is associated with distance d_{ij}. There are $p \in P$ types of products, defined by their occupied space $Spacep$ in a vehicle. Node $i \in N\backslash\{1\}$ (all customers except the depot) have demand Dip for product type $p \in P$. M is the set of transportation modes, each vehicle $m \in M$ has their capacity $Capm$, fixed cost Fm, total operating cost $cmdijm$, velocity vm and service time sim at node $i \in N\backslash\{1\}$. Whenever a vehicle is triggered, it will use a fixed amount of money for maintenance and capital amortization costs, no matter how far it goes. This is called fixed cost Fm. Operating cost is the sum of fuel and oil on average consumed by a vehicle. The total operating cost of a vehicle on route (i,j) equaled to operating cost per kilometer traveled cm multiplied with the distanced travelled on route (i,j) by vehicle m $dijm$. The time window for all nodes $i \in N\backslash\{1\}$ are [e, l] with e is the earliest time and l is the latest time vehicles can enter that node. Time window for depot is [0, L], all trucks can start leaving the depot at 0 and must be back by L. The parameters and decision variables of the proposed model are presented in Tables 1 and 2.

Table 1. Model's parameters

Parameter	Description
D_{ip}	The demand for customer node i of product type p
$Space_p$	The space of product type p
d_{ij}	Distance between node i and node j
F_m	Fixed cost when using vehicle type m
C_m	Operating cost per kilometer travelled of vehicle type m
Cap_m	Capacity of vehicle type m
v_m	Velocity of vehicle type m
S_{im}	Serving time at customer node i of vehicle m
e	Earliest time to arrive at customer zone
l	Latest time to arrive at customer zone
L	Latest returning time of all vehicles to the depot

Table 2. Model's decision variables

Parameter	Description
x_{ij}^m	Equal to 1 if vehicle type m travels directly from node i to node j and equal to 0 if otherwise
y_{ijmp}	Quantity of product type p that vehicle m carries when it leaves node i to service node j
a_{im}	Arriving time of vehicle type m at node i
l_{im}	Leaving time of vehicle type m at node i

As shown in equation (1), the model's objective function focuses on minimizing total transportation cost, including a fixed cost (F_m) and operating cost $(C_m d_{ij})$. Each cost is then multiplied to variable x_{ij}^m to make sure that only cost of triggered vehicle will be summarized. The decision variable x_{1j}^m means that only vehicle started from the depot (node $i = 1$) by vehicle m to any customer (node j) will be counted with fixed cost since this is where all vehicles start.

$$\textbf{Minimize } \sum_{m\in M}\sum_{j\in N} F_m\, x_{1j}^m + \sum_{m\in M}\sum_{\substack{i,j\in N\\ i\neq j}} C_m d_{ij}\, x_{ij}^m \quad (1)$$

$$\sum_{m\in M}\sum_{i\in N} x_{ij}^m \;, \forall j \in N\backslash\{1\} \quad (2)$$

$$\sum_{m\in M}\sum_{j\in N} x_{ij}^m \;, \forall i \in N\backslash\{1\} \quad (3)$$

$$\sum_{\substack{i\in N\\ i\neq j}} x_{ij}^m = \sum_{\substack{k\in N\\ k\neq j}} x_{jk}^m , \forall m \in M, j \in N\backslash\{1\} \quad (4)$$

Constraints (2), (3) and (4) makes sure that each customer is visited only once and that if a vehicle visits a customer, it must also depart from it. Each vehicle can be used no more than once is imposed by constraints (5) and (6).

$$\sum_{j\in N\backslash\{1\}} x_{1j}^m \le 1,\ \forall m \in M, i \in N \quad (5)$$

$$\sum_{i\in N\backslash\{1\}} x_{i1}^m \le 1,\ \forall m \in M, j \in N \quad (6)$$

Constraint (7) makes sure that there will be no vehicle assigned to route (i,j) when node i and j are the same.

$$\sum_{m\in M} x_{ij}^m = 0,\ \forall (i,j) \in N, i = j \quad (7)$$

Constraint (8) is the commodity flow constraints: they specify that the difference between the quantity of goods a vehicle carries before and after visiting a customer is equal to the demand of that customer and constraints (9) ensure that the vehicle capacity is never exceeded.

$$\sum_{m\in M}\sum_{\substack{i\in N\\ i\neq j}} y_{ijmp} - \sum_{m\in M}\sum_{\substack{k\in N\\ k\neq j}} y_{jkmp} = D_{ip}, \forall p \in P, j \in N\backslash\{1\} \quad (8)$$

$$\sum_{p\in P} D_{ip}\, Space_p x_{ij}^m \le \sum_{p\in P} y_{ijmp} Space_p \le \left(Cap_m - \sum_{p\in P} D_{ip} Space_p\right) x_{ij}^m, \forall (i,j) \in N, i \neq j, m \in M \quad (9)$$

Constraint (9) prevent any product is taken back to the depot while constraint (10) guarantees that no products will be assigned if a vehicle is not activated.

$$y_{i1mp} = 0, \forall i \in N, m \in M, p \in P \quad (10)$$

$$\sum_{p\in P} y_{ijmp} \le x_{ij}^m \times BigM, \forall (i,j) \in N, m \in M \quad (11)$$

Constraints (12) make sure that if customer j follows customer i in the route, the arrival time at customer j is equal to the departure time from customer i plus the travel time between these two customers.

$$\left(a_{jm} - \left(l_{im} + \frac{d_{ij}}{v_m}\right)\right) x_{ij}^m = 0, \forall (i,j) \in N, i \neq j, m \in M \quad (12)$$

Constraints (13) relate arrival time, departure time with service time and guarantee that their relationships are compatible to the time window. Constraints (14) and (15) are the time window for the customers and depot. Finally, constraints (16), (17), (18) and (19) restrict the values of the variables.

$$\left(l_{im} - (a_{im} + s_{im})\right)x_{ij}^{m} = 0, \forall i \in N\backslash\{1\}, j \in N, i \neq j, m \in M \quad (13)$$

$$a_{jm} \geq e \sum_{j \in N} x_{ij}^{m},\ \forall i \in N\backslash\{1\}, m \in M \quad (14)$$

$$a_{jm} \leq l \sum_{j \in N} x_{ij}^{m},\ \forall i \in N\backslash\{1\}, m \in M \quad (15)$$

$$x_{ij}^{m} \in \{0,1\},\ \forall (i,j) \in N, m \in M \quad (16)$$

$$y_{ijmp} \geq 0,\ \forall (i,j) \in N, m \in M, p \in P \quad (17)$$

$$a_{im} \geq 0,\ \forall i \in N, m \in M \quad (18)$$

$$l_{im} \geq 0,\ \forall i \in N, m \in M \quad (19)$$

In order to linearization the non-linear constraints (12) and (13), the BigM constraints method was applied and formulation the alternative constraints (20), (21), (22) and (23) below.

$$a_{jm} \geq\ l_{im} + \frac{d_{ij}}{v_m} - \left(1 - x_{ij}^{m}\right)BigM, \forall m \in M, (i,j) \in N, i \neq j \quad (20)$$

$$a_{jm} \leq\ l_{im} + \frac{d_{ij}}{v_m} + \left(1 - x_{ij}^{m}\right)BigM, \forall m \in M, (i,j) \in N, i \neq j \quad (21)$$

$$a_{im} \geq\ l_{im} - s_{im}\ - \left(1 - x_{ij}^{m}\right)BigM, \forall m \in M, i \in N\backslash\{1\}, j \in N, i \neq j \quad (22)$$

$$a_{im} \leq\ l_{im} - s_{im} + \left(1 - x_{ij}^{m}\right)BigM, \forall m \in M, i \in N\backslash\{1\}, j \in N, i \neq j \quad (23)$$

4. Case study and numerical results

DKSH Vietnam Co., Ltd. are the leading Market Expansion Services provider for companies who want to grow their business in Vietnam. Serving their business partners through their extensive global networks and industry expertise, DKSH Vietnam help companies to grow their businesses in new and existing markets. As one of Vietnam's leading business organizations, their network of 22 business locations, including offices, distribution centers, and cross-docks, across the country and 4,500 employees make them an international "local" company. DKSH Vietnam is committed to building long-term business partnerships in Vietnam. DKSH Vietnam constantly strives to bring world-class standards to industries in Vietnam while also contributing to the development of local communities.

In Healthcare branch, DKSH Vietnam has three warehouses located in Ha Noi, Da Nang and Binh Duong, and a distribution center (DC) in Can Tho. All oversea products are shipped to Binh Duong warehouse from Cat Lai seaport. Then, products will be transported to Ha Noi and Da Nang warehouses depending on the customers' location. The warehouse main functions are received products, pick and pack according to customers order, release bills and ship to customers. The Can Tho DC main responsibility is only to receive packages in large containers from Binh Duong warehouse and distribute packages into smaller truck and ship to customers. The Da Nang warehouse covers from Quang Binh to Binh Thuan. From Quang Binh up to the north is covered by Ha Noi warehouse, and from Binh Thuan province down to the south is cover by Binh Duong warehouse.

Right now, DKSH Vietnam delivery policy is that one truck is responsible for one province's order. The capacity of the truck is depended on the province's demand. Trucks will get packages from warehouses and transport to the province's city. From there, a local transportation team hired by DKSH Vietnam consisting of 3-5 people, depending on the province size, will receive packages from the truck and shipped to customers. After distributing packages, the truck will ship to customers living in the city only and come back to warehouses after finish shipping. The latest time for shipping is 6 PM then all trucks must go back and be available at the depot before 1 AM of the next day. Customer orders cut off time is 6 PM each day. Orders are then sent to warehouses, where they will be processed, pick and pack into packages, and loaded on trucks. Only after the cut off time will DKSH Vietnam know exactly each customer demand for the next delivery and each truck delivery schedule will be generated. By 2 AM the next morning, trucks are loaded with packages and the transportation cycle begins again. Each province or city is divided into many areas with different delivery schedule. Delivery frequency ranges from 1 to 6 days a week, depending on customer need and distance from the warehouse. Usually, big city centers will be delivered 6 days a week. DKSH Vietnam continues to expand its market by signing contracts with new clients each year. In 2018, Sanofi became DKSH Vietnam's new client and is one of the biggest clients. Its sales last year accounted for 28% of total sales. DKSH Vietnam classified their customers into 4 groups as Table 3.

Table 3. DKSH Vietnam customer classification

Group	Average value per order in a month (VND)	Proportion
A	Over 15 million	5%
B	10 – 15 million	12%
C	5 – 10 million	28%
D	Under 5 million	55%

As can be seen, more than half of customers usually place orders under 5 million VND and the average SKU per order is 5. This means that each time customers place an order they usually buy few products and spend little money on it.

4.1. Case study data

Customers demand data collected from DKSH Vietnam is processed by grouping customer in some districts in a province into one node. As can be seen, some nodes are different areas of one province. Unlike DKSH Vietnam transportation rule, where they decided that one vehicle is responsible for one province, in this model, the vehicle can transport freely between nodes that are not in one province. One month demands in 2018 of these nodes were chosen randomly and the average is used for the model customers demand input as Table 4.

Table 4. Node demand

Node	SKU				
	1	2	3	4	5
1	0	0	0	0	0
2	14	8	11	5	2
3	10	13	5	9	1
4	5	6	3	2	2
5	3	2	7	5	0
6	6	5	2	5	4
7	9	11	5	7	2
8	3	5	5	2	0
9	5	2	2	3	1
10	6	5	2	2	1
11	7	4	4	1	2
12	7	3	2	1	1
13	6	4	2	1	0
14	4	3	5	1	0
15	6	2	0	1	3
16	7	1	5	2	4
17	9	3	0	2	0
18	12	2	0	0	2
19	5	7	2	2	1
20	4	1	0	5	4
21	3	1	1	2	1
22	14	2	0	3	0
23	6	2	1	4	0
24	4	6	3	10	4
25	7	5	2	9	2
26	5	7	9	0	3
27	11	9	5	3	2
28	2	1	3	0	5
29	9	1	1	2	0
30	11	2	3	7	4
31	14	2	1	0	0
32	9	2	3	1	0
33	8	3	2	3	2
34	12	4	3	0	0
35	14	3	2	0	1
36	5	2	5	3	0
37	10	3	4	0	0
38	9	4	0	5	1
39	7	3	2	1	0
40	7	3	0	3	0
41	8	2	1	0	1
42	6	4	0	3	0
43	5	2	3	1	0
44	9	2	3	0	0
45	7	5	1	2	1

Each node demand contains many packages of different size, each package is for a customer as Table 5. The distance between these nodes is estimated based on the position where the truck drops off packages to a local transportation team and are calculated by using Python linked with Google map.

Table 5. Product space requirement

SKUs	1	2	3	4	5
Space (m^3)	0.1	0.2	0.3	0.4	0.5

All vehicles will start leaving the depot at 2 AM each day. This is considered as time 0 in the model when all transportation begins. Each vehicle type has different capacity, velocity. To create equality between vehicles with the same type, their parameters (capacity, velocity, operating cost per km travelled, service time for each customer will be exactly the same. This means, all trucks of the same type will have equal probability chosen by the model. Vehicle velocity may varies depended on the size of the road. So, their average velocity will be taken, which is 40 – 45km/h, depending on the truck size. Bigger truck will have slower speed. These data are consolidated by DKSH Vietnam given data as well as reference by sShip transportation company trucking fee as Table 6. Meanwhile the time window for customer ($[e, l]$) is 06:00-14:00 with the latest depot depart time (L) is 01:00 every day.

Table 6. Vehicle characteristics

Vehicle ID	1-10	11-20	21-23
Vehicle type (ton)	2	2.5	15
Capacity (m3)	13	16	66
Velocity (km/h)	45	45	40
Fixed cost (VND)	210,000	230,000	350,000
Operating cost (VND/km)	21,500	22,500	31,500
Service time (hour)	0.3	0.4	0.5

4.2. Optimization results

In Table 7, the optimization results are presented, which proposed the total transportation cost is 611,174,000 VND (~27,407 USD) with 16 vehicles are triggered.

Table 7. Vehicle routing optimization result

Vehicle ID	Route Node Sequence
1	Depot-2-19-Depot
2	Depot-4-5-6-Depot
3	Depot-25-39-43-45-Depot
4	Depot-37-38-Depot
5	Depot-21-44-Depot
6	Depot-11-10-Depot
7	Depot-7-27-Depot
8	Depot-29-26-28-30-Depot
9	Depot-16-Depot
10	Depot-40-Depot
11	Depot-34-3-Depot
12	Depot-24-35-20-Depot
14	Depot-15-14-32-31-Depot
15	Depot-22-23- Depot
18	Depot-33-12-13-42-Depot
20	Depot-8-9-36-17-18-41-Depot

These results are compared with the cost of routes actually covered by the trucks of the company at the same time as Figure 1. In both cases, the route cost is calculated by adding the fixed and operating costs.

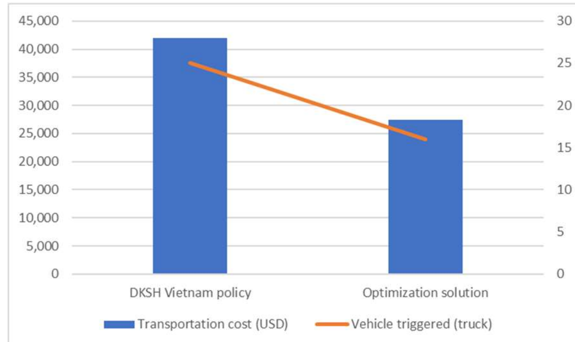


Figure 1. Transportation performance

As state before, there are 45 nodes comes from 25 provinces. With DKSH Vietnam transportation rule, which is one vehicle responsible for one province, there are 25 triggered vehicles and it costs them around 42,000 USD transportation cost. This does not even cover the transportation cost when the trucks go around the province's city for delivery. Comparing both models, it is clear that with the new model, DKSH Vietnam can reduces 9 trucks, saving around 14,500USD per day, or 3,253,480USD per year.

The advantage of this proposal is that each vehicle does not strictly need to be visited only one province. By dividing provinces into one or many nodes based on their demands, a vehicle can flexibly travel between two nodes of, maybe, different provinces, as long as this route saves more cost. Another advantage of this is that instead of using a big truck to deliver to customers in cities, the truck will now just drop off packages and one or two local transportation employees, depend on how many customer orders, will deliver them to customers. By switching from a truck to motorbikes, packages will be delivered faster, considering Vietnam traffic characteristics. Furthermore, it will save more transportation cost since a 2-ton truck cost 21,500VND/km while two motorbikes cost only 10,000VND/km (5,000VND/km each). For a more practical and out of scope comparison, transportation employee costs will also now be included. There are two types of transportation employee, local transportation employee, and on-truck transportation employee. Local transportation employees are DKSH Vietnam employees who based in each province with their duty to receive a package from a truck and distribute it to customers. On-truck transportation employees are the ones who go with the truck to help the driver distributing packages to local employees and deliver packages to the city's customers while the driver stays with the truck. With the new method, the on-truck employees are not needed anymore, since now all trucks do not deliver directly to any customer. In this particular situation, 25 of on-truck employees are a shift to 25-35 local employees. Salaries for these two types of employees are 6 million VND/person/month. With about 10 people differences, it will cost around 60 million VND/month more (~2,690USD). However, this is not much comparing to the 14,500USD/day saving in transportation cost.

5. Conclusion

This paper proposed a mathematical model to solve a real-life transportation problem in a pharmaceutical distribution company, DKSH Vietnam. The problem is classified as a heterogeneous fleet vehicle routing problem with time windows problem, which was modified to a real-life situation and introduced a linearization method for some constraints.

The proposed approach has fulfilled the research objective, which is to minimize total transportation cost by determining the optimal set of routes for a fleet of vehicles. It has offered a better solution in terms of distribution cost and a number of trucks used compared to the current practice while at the same time still respect all problem constraints and time window. DKSH Vietnam could apply this model to reality easily, save more transportation and system cost than what they are doing right now.

Most of the problem aspects have been covered in this model, however, there are still some limitations due to solving time constraint. For future work, the solving time should be reduced to at most 30 minutes to reach optimal solution. In order to do that, other solving method, like heuristics, could be developed. The model could be improved to be applied for more complex problems at Southern region, where a DC is involved, and Middle region since there are longer distance and the lead time varies from 1 to 2 days. For a more practical and complete solution, other transportation factors such as toll, rush hour, reducing emission gas should also be taken into consideration. The number of product types could increase more to be more realistic.

References

Abdullahi, H., Reyes-Rubiano, L., Ouelhadj, D., Faulin, J., & Juan, A. A. (2021). Modelling and multi-criteria analysis of the sustainability dimensions for the green vehicle routing problem. *European Journal of Operational Research, 292*(1), 143-154. doi:10.1016/j.ejor.2020.10.028

Chen, D., Pan, S., Chen, Q., & Liu, J. (2020). Vehicle routing problem of contactless joint distribution service during COVID-19 pandemic. *Transp Res Interdiscip Perspect, 8*, 100233. doi:10.1016/j.trip.2020.100233

Elshaer, R., & Awad, H. (2020). A taxonomic review of metaheuristic algorithms for solving the vehicle routing problem and its variants. *Computers & Industrial Engineering, 140*, 106242. doi:10.1016/j.cie.2019.106242

Hsu, C.-I., Hung, S.-F., & Li, H.-C. (2007). Vehicle routing problem with time-windows for perishable food delivery. *Journal of Food Engineering, 80*(2), 465-475. doi:10.1016/j.jfoodeng.2006.05.029

Li, H., Wang, H., Chen, J., & Bai, M. (2020). Two-echelon vehicle routing problem with time windows and mobile satellites. *Transportation Research Part B: Methodological, 138*, 179-201. doi:10.1016/j.trb.2020.05.010

Neira, D. A., Aguayo, M. M., De la Fuente, R., & Klapp, M. A. (2020). New compact integer programming formulations for the multi-trip vehicle routing problem with time windows. *Computers & Industrial Engineering, 144*, 106399. doi:10.1016/j.cie.2020.106399

Öztop, H., Kizilay, D., & Çil, Z. A. (2020). Mathematical models for the periodic vehicle routing problem with time windows and time spread constraints. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA), 11*(1), 10-23. doi:10.11121/ijocta.01.2021.00899

Pan, B., Zhang, Z., & Lim, A. (2021). Multi-trip time-dependent vehicle routing problem with time windows. *European Journal of Operational Research, 291*(1), 218-231. doi:10.1016/j.ejor.2020.09.022

Redi, A. A. N. P., Jewpanya, P., Kurniawan, A. C., Persada, S. F., Nadlifatin, R., & Dewi, O. A. C. (2020). A Simulated Annealing Algorithm for Solving Two-Echelon Vehicle Routing Problem with Locker Facilities. *Algorithms, 13*(9), 218. doi:10.3390/a13090218

Vidal, T., Laporte, G., & Matl, P. (2020). A concise guide to existing and emerging vehicle routing problem variants. *European Journal of Operational Research, 286*(2), 401-416. doi:10.1016/j.ejor.2019.10.010

Zhang, S., Chen, M., Zhang, W., & Zhuang, X. (2020). Fuzzy optimization model for electric vehicle routing problem with time windows and recharging stations. *Expert Systems with Applications, 145*, 113123. doi:10.1016/j.eswa.2019.113123

ORDER PICKING OPTIMIZATION MATHEMATICAL MODEL WITH ORDER BATCHING APPROACH: A CASE STUDY OF VIETNAM THIRD-PARTY LOGISTICS COMPANY

Nguyen Van Chung[1], Le Nguyen Tung[1] and Luong-Nhat Nhieu[2*]

[1]International University – Vietnam National University HCMC, Hochiminh City 70000,

nvchung@hcmiu.edu.vn (N.V.C)

lenguyentung11@gmail.com (L.N.T)

[2]National Kaohsiung University of Science and Technology, Kaohsiung 80778,*

nnluong.iem@gmail.com (L.-N.N.)

Abstract: In warehouse operation, orders with numerous product variants in small volume are often received and asked for on-time delivery. However, the picking process becomes very wasteful and costly when the pickers handle these small orders separately. In this paper, a mathematical model is proposed to assist grouping orders into a batch, then the distance to pick orders will be calculated, before and after using the model. Through the proposed model, all received orders will be divided into batches to minimize the distance travel. To prove the realistic of the model, a case in the YCH company is presented, and the outcomes from the mathematical model provide a better situation, which the order picking process can be better arranged.

Keywords: Order Picking problem, Order Batching, Mathematical model, Third-party logistics.

1. Introduction

A distinction is made between warehouses and distribution centers for articles that are stored and retrieved to fulfill customer orders. The variety of storage units in the warehouse and the quantity in which the items are requested can be influencing factors on the storage system. There are many things to consider with internal and external storage systems, and there are several factors that affect the storage systems. While it may appear that the only function of a warehouse is warehousing, picking plays an important role in reducing costs and improving productivity in warehouse and distribution center operations. Defined as the process of providing warehouse products in response to a specific customer requirement, picking is the most labor-intensive process in the warehouse with both manual systems and automated systems. Therefore, order picking is a top priority to improve productivity.

The appearance of the order stacking not only increases the picking efficiency, but also allocates the minimum storage resources and fulfills customer requests within a certain period of time. Since articles from two or more orders are combined into one picking trip during order stacking, this process takes place before the actual picking trip. This process helps increase efficiency, as the order picker can drive and pick the same items for several orders at the same time. With the order stacking strategy, employees can save time and travel if similar item locations can be processed at the same time.

However, the problem of allocating an appropriate number of resources to select the jobs is ignored. Processing small orders separately can lead to higher costs as order picking is a labor-intensive and costly activity. For example, if the order picker starts the order picking tour for each individual order with only one or two storage units in small quantities, the effectiveness and efficiency are low, since the capacity of each work step cannot be fully utilized. Additionally, due to limited resources in the warehouse, it may not be possible to fulfill all orders on time. For this reason, to control costs and efficiency, it is crucial to make the picking process less labor-intensive. In order for the picking process to be carried out effectively, the equipment that is to be assigned for picking the batch job must be specified. YCH's warehouse imports items from manufacturers and then exports the products to various forms of business such as retail agents, grocery stores, and supermarkets. There are different types of products and different quantities that need to be delivered to businesses and retailers. The most serious and common problem is that the picking process wastes a lot of time in warehouse operations. There are many reasons for this situation, the climax of which is that orders with many product variants are often received in small quantities and requested for on-time delivery. The order pickers often have the order picking tour for each individual order with only one or two storage units. The quantity is not large, but there are so many separate orders that the picking process takes time, the effectiveness becomes low as the ability of each tour to pick items cannot be exploited. Moreover, with the limited human resources, the available staff may not be able to fulfill all orders on time. The picking process is a labor-intensive and costly activity that would severely impact operations if order pickers were to handle small orders separately. The company's expectation to solve this problem is to manage the picking process by grouping orders as a batch in a single picking trip to minimize travel distance and save time to complete the customer order, make the picking process less cumbersome and improve efficiency. The goals of this paper are to model and minimize the path to picking the orders.

2. Literature review

The subject of warehouse picking issues (OPP) has been considered by many researchers over the past year. Many articles have been published based on real life questions. Several authors presented a literature search and conducted a survey on the warehouses or distribution centers. Matson et al. (1982) gave a general overview of material flow studies on a variety of topics such as robotics, material handling, transfer lines, flexible manufacturing systems (FMS), and device choice and models for warehousing.

Ashayeriet al. (1985) dealt specifically with warehousing and collected 9 pieces of data on systematic approaches and simulation procedures for the bottlenecks of warehouse design optimization. Cormier, et al. (1992) reviewed the more modern literature dealing with optimizing bearing design and operation. Authors mentioned three warehouse models: throughput capacity models, warehouse capacity models, and warehouse design models. Throughput capacity models concentrate on the guidelines for picking, batch processing, storage, and allocation. Jeroen (1999) presented a literature review on methods and techniques for planning and controlling storage systems, even if they are not easy to solve. The author focused on the problem of bin allocation on a tactical level and identified three problems of order stacking, routing and dwell positioning as control problems. He pointed out that numerous batching heuristics have been presented in the literature to minimize travel time, and most of the heuristics adopted seed order and proximity batching methods.

Batching is a well-known and efficient strategy to diminish the average travel time per job and thus improve system performance. A batch is a sequence of orders that are picked in a single tour. The total of the positions of the orders in the batch must not exceed the storage capability of the picking vehicle. Additionally, we can maximize system throughput by generating large quantities of orders at nearby picking locations.

Additionally, selecting orders only from nearby picking locations can unduly delay orders down the aisle in the warehouse. As Gademann and Van de Velde (2005) mentioned in their work “Order Batching to Minimize Total Travel Time in a Parallel-Aisles Warehouse”, the problem of order stacking relates to the decision to effectively batch many orders to improve and optimize picking processes. If the customer orders contain a large assortment of articles, small order sizes and short response times, order stacking applies. By batching many orders prior to the picking process and processing a group of orders at the same time, overall picking time and efficiency can be significantly improved (Peterson, 2000). In general, travel times and distances of the picker are the crucial elements in increasing the picking efficiency. Order batching, which refers to the actions of picking two or more requests together in one trip, is a common practice in order picking. If orders in the same picking areas are assembled and collected in the same order picking trip, order batching can decrease the absolute request picking travel separate.

Several different order batching strategies have been introduced by research by Y.C.Ho and Y.Y. Tseng (2006). One of them executes based on a selection rule while the other relies on instructions to select a seed-order. The difference of the seed-order selection strategy is to select the first order for each batch and then add the other suitable orders to the batch. In the study of Cathy et al. (2014), a fuzzy optimization mathematical model was developed to classify orders into different batches and determine the picking sequence of these batches. The application’s results of this method in the management practice of a logistics warehouse have shown that the order picking process can be organized more efficiently.

To support the problem of order batching, one linear and another non-linear optimal mathematical model was introduced in Vinod's publication (1979). The order batching process in these models takes on attribute sequences of orders. Meanwhile, dummy orders are used in order batching with fixed-size batches in the mathematical model of Armstrong et al (1979). In addition, integer and quadratic programming models for the order batching problem have also been enhanced by later studies (Kusiak 1986).

Rosenwein (1996) presented and compared various heuristics for order stacking in multi-gas storage systems. The author used some metrics (the Minimum Distance Metric for Additional Aisle (MAA) and Center of Gravity (COG)) that approximated the relative proximity of an order pair.

Elsayed, et al. (1989) compared four heuristics under the hypothesis that the overall number of orders is normally distributed, while the overall number of entries in an order and the quantity of each item are uniformly distributed random variables. They demonstrated that the best algorithm is to first sort each order large or small based on a predetermined proportion of the vehicle's capacity. Chew and Tang (1999) also investigate a discrete picking system in a rectangular warehouse. Her focus is on studying the implications of a lot size in a real-time system. Their probabilistic model determines the number of items per order (the lot size) and then develops a queuing model in which incoming orders are queued and stacked before picking. Two very common procedures of collecting orders in a warehouse are picking (OP) and zone picking (ZP) for a picking process (Heragu, 1997). In order picking, an operator or order picking vehicle is responsible for picking all items in a batch or route, while in zone picking an operator assigned to a zone (several aisles) is responsible for picking all items in orders in a batch. With zone picking, the order picker must drive within a zone. The aim of order picking is to organize the locations of the items to be visited in order to minimize the travel time of an order picking vehicle. It can eliminate unnecessary time in the picking process.

Goetschaleks et al. (1988) presented an efficient method for order picking in a warehouse with an aisle width that should not be neglected. Two-way traffic is possible in wide aisles, traffic can turn and pass, and lifting with forklifts is possible. They proposed that over 30% of the savings can be made by picking both sides of the aisle instead of picking one side first and then the other. Then Goetschaleks et al. (1989) dealt with the problem of determining the optimal stopping position of an order picking vehicle in an alley when the order picker can perform several picks per stop. They suggested an efficient dynamic programming algorithm in case the picker's travel time is measured with the linear metric.

Gudehus (1973) described the widespread band heuristic. This heuristic divides the shelf of the warehouse into two horizontal bands or areas. First, the locations on the lower band are visited with increasing x-coordinates; then the locations on the upper band are visited with decreasing x-coordinates. The authors claim that any even number of ribbons can be used. Bozer (1985) derived an analytical expression for the expression for the expected tour length of the band heuristic and the optimal number of bands as a function of the number of picks.

Kanet et al. (1986) proposed a mixed non-linear programming formulation for the problem of choosing from alternative picking locations in order to minimize a combination of error costs and fixed and variable picking costs. The authors assumed that the variable costs are a function of travel time, while the fixed picking costs depend on things like the loading and unloading times of the pallets.

3. Methodology

There are five main steps that should be considered in the process of finding and solving OPP as shown in Figure 1.

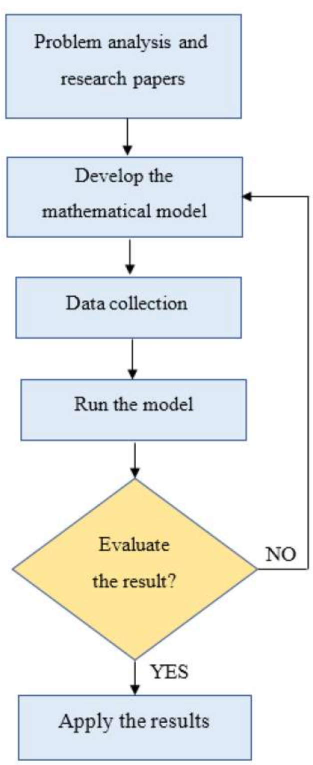


Figure 1. Research methodology

In first step, to point out solution and improve the problem for the company, the first and the most importance thing is understanding about the company. The overview of the company such as its network, working environment and policies is need to understand. Besides, we want to know how the warehouse operates, how the order picking process works and what kind of method the company apply for warehouse. To solve the above problem, the necessary information was observed and memorized. Moreover, the staff in-depth interview conducted to get more information than observed information. In the model development step, a mathematical model was developed with the objective is to minimize the total travel distance to pick all orders and its constraints needed to be found base on real issue. In the next step, the model data, which is mentioned in section 3.2, will be collection in primary and secondary forms. Then the IBM CPLEX Solver is using to finding the optimal solutions. These solutions are then evaluated by the actual operators of the company leading to decisions to improve the model if necessary. Finally, the optimal results and the newly developed model are synthesized and documented to support the company's operational decisions.

3.1. Develop the mathematical model

The goal of this model is to minimize the total distance traveled in order picking process and the constraints that need to be found based on a real-world problem. Basing on these requirements, using CPLEX to build the model and find the optimal solution. The language interface base on the C interface and it can connect to Excel or MATLAB, so get familiar with CPLEX programming is necessary. This study uses a combination of two serial optimization models including the well-known Traveling Salesman Problem (TSP) model, which is applied to calculate the distance of batches and a proposed model, which presented below, to solve the problem and its input is results of the first model. Besides that, there are three matrices that help in the process of defining batches and the travel distance. The first matrix is matrix U, which show all the batches results. The second matrix is matrix nU, each row of this matrix is the total elements of the corresponding row in matrix U. The number of rows of the U and nU matrices is determined by formula (1). The third matrix is matrix

R, each row of this matrix shows a result of batching method.

$$\sum_{p=1}^{M} \frac{N!}{(N-p)!p!} \quad (1)$$

The proposed model's parameters variable was shown in Table 1. These parameters describe the number of orders received by the system; the number of batches generated during the order batching process. In addition, the parameters of the capacity of the system and each order are also considered.

Table 1. Mathematical model' parameters

Notation	Description
N	Number of orders received
T	Number of batches created
M	Maximum number of orders in one batch, M = N – T +1
K	Number of ways to divide N orders into T batches (K is also the number of rows in matrix R)
F_c	The picking facility capacity
F_w	The picking volume capacity
F_V	The picking weight capacity
Q_i	Capacity of order i
W_i	Weight of order i
V_i	Volume of order i
D_k	The optimal distance when collecting products of orders in k^{th} batch.

Equation (2) describes the model's objective function as minimizing the total order picking distance. In which, the decision variable x_k is equal to 1 if the k^{th} combination is chosen and equal to 0 otherwise.

$$\boldsymbol{Minimize}\ Z = \sum_{k=1}^{K}\left[x_k \sum_{b=1}^{T} D\big(R(k,b)\big)\right] \quad (2)$$

The constraint (3) ensures that the total number of the same-batch picked items must not exceed the capacity of the system.

$$\left(\sum_{i=1}^{nU_{R(k,b)}} Q_{U(R(k,b),i)}\right) x_k \leq F_c, \forall k \in (1, \dots, K), \forall b \in (1, \dots, M) \quad (3)$$

Meanwhile, the constraint (4) limits the total batch weight of picked items. The constraint (5) limits the total batch volume of picked items.

$$\left(\sum_{i=1}^{nU_{R(k,b)}} W_{U(R(k,b),i)}\right) x_k \leq F_W, \forall k \in (1, \dots, K), \forall b \in (1, \dots, M) \quad (4)$$

$$\left(\sum_{i=1}^{nU_{R(k,b)}} V_{U(R(k,b),i)}\right) x_k \leq F_V, \forall k \in (1, \dots, K), \forall b \in (1, \dots, M) \quad (5)$$

Lastly, the constraint (5) states that there is only one way to assign N orders into T batches.

$$\sum_{k=1}^{K} x_k = 1 \quad (6)$$

3.2. Data structure

The data to be collected for these two models includes two groups: warehouse layout data and customer order data. warehouse layout data should include information such as the area of warehouse, number of aisles and number of storage locations in each aisle, the width of each storage space. While, customer order data provides information about delivery date, item name and code with location, weight and volume.

4. Numerical results

The YCH Vietnam office opened in 2009 to capitalize on the burgeoning economic activity of the ASEAN countries. With around 499 employees, YCH has two headquarters in Ha Noi and Binh Duong provinces. YCH Protrade DistriPark in Binh Duong Province is one of the most modern warehouses in Vietnam, strategically located behind Song Than Industrial Park and Vietnam Singapore Industrial Park and only 17 km from Tan Son Nhat International Airport. With an ever-growing presence in Hanoi and Binh Duong, YCH Vietnam serves some of the fastest growing F&B companies and seeks to leverage its extensive freight capabilities to deliver effective results to our world class customers looking to expand in the emerging region. Current layout of YCH company warehouse was shown in Figure 10. The model is applied to the order picking process with four sample data sets with different orders as shown in Tables 1, 2, 3, and 4. As in dataset 1, this scenario is intended to test the model's effectiveness with a group of five orders with capacity, volume, volume, and item position within the warehouse layout. For example, the first order of dataset 1 has a total capacity, weight, and volume of 97, 49, and 0.3 respectively. This order includes three items at positions C16, D4 and F3 in the warehouse. Similarly, other scenarios are built in the form of different data sets to evaluate the performance of the model.

Table 2. Dataset 1

Order	Capacity	Weight (kg)	Volume (m3)	Location
1	97	49	0.9	C16, D4, F3
2	51	21	0.11	D2
3	35	88	0.28	A4, A8, A15, B3, D3
4	39	55	0.44	E8, E13
5	68	1	0.53	D8, D12, E15, F11, F6, F3

Table 3. Dataset 2

Order	Capacity	Weight (kg)	Volume (m3)	Location
1	36	64	0.7	A19, B3, C4, D14, E16, E5
2	13	35	0.78	A5, B17, B1, C10, D13, E6, E1, F7
3	65	51	0.23	A2, A12, A16, C15, D3
4	41	63	0.04	B10, B6, C6

Table 4. Dataset 3

Order	Capacity	Weight (kg)	Volume (m3)	Location
1	55	45	0.99	D5, D13, E16, E10, F5, F19
2	17	84	0.26	C12, D3, E17
3	51	91	0.23	E17, F13
4	33	42	0.31	E7, E16, F14
5	35	95	0.07	D16, E2, F2, F12
6	49	7	0.45	A18, C7

Table 5. Dataset 4

Order	Capacity	Weight (kg)	Volume (m3)	Location
1	74	54	0.81	A4, A11, A13, B17, B15
2	26	69	0.25	A16, B3, C6, D12, E12
3	80	11	0.12	A19, B9, C13, C16
4	90	17	0.62	C13, C18
5	27	80	0.09	A7, A19, B3, C15, D6, E14, F13
6	86	25	0.57	B9, B13
7	59	19	0.93	A5, B9, B19, C7, D10, D15

As depicted in Figures 2 and Figure 3, The total travel distance and the number of picking tours are both improved through the solutions of model optimization.

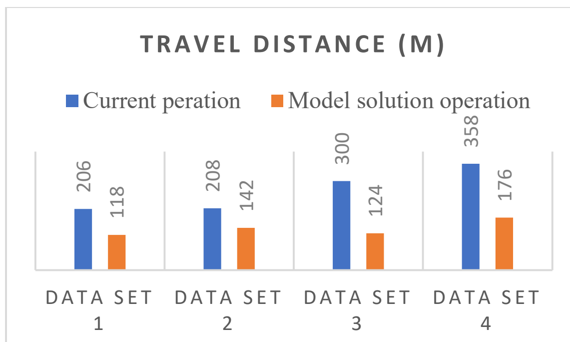


Figure 2. Methodology flow chart

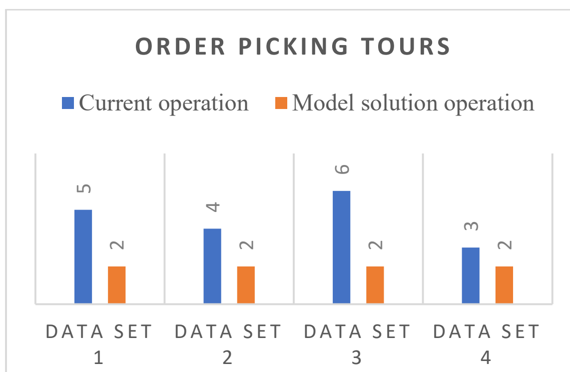


Figure 3. Methodology flow chart

The results show that both the number of trips and the total distance traveled can be improved by at least 50% and 31.73%, respectively. Through the proposed model and its results, the labor resources can be improved. If each order is picked individually, it will be labor intensive more than when all orders are grouped. By applying this method to the YCH warehouse, this saves a lot of energy for the workers here. You can increase productivity to do other types of tasks such as inspection and packing activities. As a result, the system will increase the effective use of labor resources and the waste of energy will be reduced. In addition, the use of the order stacking method reduces the total distance of the order picker. The reason for this is that the order picker does not have to return to the input / output point after completing an order. Instead, he can pick the items for multiple orders in just one tour, which could reduce the total distance traveled. In addition, orders in one batch have the same location and aisles, which is an important factor in reducing the distance for order pickers. Another advantage of the batch process is the time it saves. If the total distance traveled is reduced, the time it takes to complete the order is also reduced.

5. Conclusion

Nowadays, consumers often place the delivery orders to the warehouse according to their sales needs. The orders with a high multitude of products in small quantities for replenishment are lazily accepted. If all customer orders cannot be satisfied on time, customers suffer reduce and the warehouse also loses its profit. Therefore, the picking process and its effectiveness is important to complete orders with various requested items on time. To improve the picking process, the model is presented in this study in order to determine the batching results of orders and to minimize the total route to picking orders.

Reference

Armstrong, R.D., Cook, W.D., and Saipe, A.L., (1979). *Optimal batching in a semi-automated order picking system*, Journal of Operational Research Society.

Ashayeri, J. and Gelders, L.F., (1985). *Warehouse Design Optimization*, European Journal of Operational Research.

Bozer, Y. A., and Sharp, G. P., (1985) *An empirical evaluation of general-purpose automated order accumulation and sortation system used in batch picking*, Material Flow.

Caron, F., Marchet, G., Perego, A., (2000). *Optimal layout in low level picker-to-part systems*, International Journal of Production Research, **38**.

Cormier, G. and Gunn, E. A., (1992). *A review of warehouse models*, European Journal of Operational Research.

Cathy, H.Y. Lam., Choy, K.L., Ho, G.T.S., & Lee, C.K.M., (2014). *An order-picking operations system for managing the batching activities in a warehouse*, International Journal of System Science.

Chabot, T., Lahyani, R., Coelho, L., & Renaud, J. (2016). *Order picking problems underweight, fragility and category constraints*, International Journal of Production Research.

Goetschalckx, M., and Ratliff, H. D., (1988a) Order picking in an aisle, HE Transactions.

Goetschalckx, M., and Ratliff, H. D., (1988b). *Sequencing picking operations in a man-aboard order picking system*, Material Flow.

Kanet, J. J., and Ramirez, R. G., (1986). *Optimal stock picking decisions in automatic storage and retrieval system*, Omega International Journal of Management Science.

Le-Duc, T., De Koster, R., (2005b). *Layout optimization for classbased storage strategy warehouses. In: de Koster, R., Delfmann, W. (Eds.)*, Supply Chain Management–European Perspectives. CBS Press, Copenhagen.

Matusiak, M., R. de Koster, L. Kroon, & J. Saarinen. (2014). *A Fast Simulated Annealing Method for Batching Precedence-constrained Customer Orders in a Warehouse*. European Journal of Operational Research.

Roodbergen, K. J. (2001). *Layout and routing methods for warehouses*. Ph.D. thesis, RSM Erasmus University, the Netherlands.

Rosenwein, M. B., (1996). *A comparison of heuristics for the problem of batching orders for warehouse selection*, International Journal of Production Research.

Vinod H. D., (1979). *Integer programming and the theory of grouping*, Mathematica.

Y.A.Bozer., & J.W.Kile., (2008). *Order batching in walk-and-pick order picking systems*. International Journal of Production Research.

Y.-C. Ho & Y.-Y.Tseng., (2006). *A study on order batching methods of order-picking in a distribution center with two cross-aisles*. International Journal of Production Research

DIFFICULTIES AND PROBLEMS IN ESTABLISHING AND APPLYING ISO 9001:2015 STANDARD AT ENTERPRISES IN VIETNAM

Ho Thi Phuong Thao[1] and Le Huong Duong[2]

[1] Faculty of Commerce, Van Lang University, Address: 69/68 Dang Thuy Tram Street, Ward 13, Binh Thanh District, Ho Chi Minh City, Vietnam, Email: thao.htp@vlu.edu.vn

[2] Faculty of Commerce, Van Lang University, Address: 69/68 Dang Thuy Tram Street, Ward 13, Binh Thanh District, Ho Chi Minh City, Vietnam, Email: duong.lh@vlu.edu.vn

Abstract: This paper surveys the difficulties and problems in establishing and applying ISO 9001 standardat enterprises located in Vietnam. Using the method of case studies, this study identifies problem groups according to the classification of ISO 9001. The survey of 169 enterprises in Vietnam showed that while ISO 9001 certification has a positive impact on the performance of enterprises. Some of the problems and entanglements encountered in certification include failures in setting the purpose of application, time for setting up the quality management system, leadership and commitment, resources and training activities, awareness and engagement of people, documented information system establishment, application, taking corrective actions, monitoring, measuring and evaluation. Moreover, this paper also suggests solutions to the above difficulties and problems.

Keywords: ISO 9001; Applying Quality System; Difficulty/ problem in applying ISO 9001

1. Introduction

The ISO 9000 standard is standardized for a quality management system and is becoming increasingly popular worldwide. The ISO 9000 family was first issued by the ISO in 1987 and reviewed in 1994 (2nd version), 2000 (3rd version), 2008 (4th version), 2015 (lastest version). So, was the increase in the number of certifications that had come from the perception of the benefits of certification? Did ISO 9000 help improve business results? (Starke et al, 2012).

In terms of the size of the economy and the number of businesses, the number of ISO 9000 certificates between the US and Europe should be equal but in reality businesses in the US did not seem to appreciate ISO 9000 certification as in Europe (Martinez-Costa and Martinez-Lorente, 2007). In 2020, the number of US ISO 9000 certification was less than France and not as half as of Spain. Some research discovered that the reason why some enterprises pursued certification was the requirement from customers. Besides the two-digit growth rated over the past decade, the export-oriented production direction prompted businesses in the country to pursue certification to open their doors for products entering the European market (Martinez-Costa and Martinez-Lorente, 2007).

The actual figures showed that the number of certifications in the world increased continuously over time, but the true benefits of certification have remained a matter of debate. In some cases, the results of the study were very different and even contradictory. One of the causes of this difference was due to research using very different approaches (Casadesus, 2005). Despite this, several case research confirmed that ISO 9000 certification could bring businesses certain benefits (Beattie and Sohal, 1999; Douglas et al, 1999; Withers and Ebrahimpour, 2000).

Although ISO 9000 certification were pursued and implemented globally, there was little research that explored the problems and difficulties of establishing and applying ISO 9000 certification to operational results by case research. The data gathered from case research was an ideal method to describe success stories or practices of ISO 9000 certification. However, there was insufficient data to justify the relationship between ISO 9000 certification and business efficiency (Heras et al, 2002).

The ISO 9001, Quality Management System - Requirements, (one standard of the ISO 9000 series) has been the most widely applied by the enterprises. The currently effective edition of this standard is ISO 9001:2015. According to the survey data of ISO Standard Committee, as of December 2018, the total number of valid ISO 9001 certificates issued to enterprises worldwide is 878,664 and in Vietnam is 3774 (Department of Science and Technology - Ministry of Industry and Trade, 2020). Although many companies have been granted certificates, the companies which are setting up and applying this standard still face to the following difficulties and problems, which are mentioned in this article. Within the scope of this article, the authors present the difficulties and problems, including but not limited to the most common ones and the solutions are recommended for solving them.

2. Literature review and research development

Academic evidence suggested that there were differences in study results that were even contradictory in terms of the impact of ISO 9000 certification on business results. Some research showed that ISO 9000 helped improve operational efficiency, increase customer satisfaction, improve market share, save costs, improve competitive advantage, product quality, revenue and income (Casadesus and Gimenez, 2000; Douglas et al, 1999; Corbett et al, 2005; Dunu and Ayokanmbi, 2008). Meanwhile, some other research criticized ISO 9000 certification as costly and time-consuming but did not bring positive financial results to the business (Han et al, 2007).

In research of the impact of ISO 9000 certification, research objectives could be the same but the method of measuring variables could vary between studies (Boiral, 2012). The results of theoretical survey from 2000 to 2020 showed that there were all 45 variables, divided into four variable groups corresponding to four research topics: performance, governance results, difficulty groups, and relationship tally groups.

Research focusing on variables such as productivity, profitability, communication and corporate image, were often seen as the strengths of certification but did not care enough about the difficulties that arose during the certification process. Therefore, businesses did not fully survey the common difficulties and challenges of businesses when implementing ISO 9000. The past research should clarify the role of variables such as implicit knowledge, behaviors against change and contradictions between reason and practice (Boiral, 2012).

Other research that data were objective financial indicators of the business. Testing hypotheses with objective data was useful, but very few research used objective data to assess the impact of ISO 9000 certification on business financial results (Sharma, 2005).

Some research used qualitative data that often used factor analysis to survey the perceptions or attitudes of relevant subjects about the positive results of ISO 9000 certification or difficulties in the certification process while identifying factors that measure each type of enterprise's results (Chow-Chua et al, 2003). The data was academically acceptable but can create subject when answering questions, making the data possible to be misleading (Casadesus and Karapetrovic, 2005). This is the limitation of research that go in this approach.

3. Survey results

The article is written basing on the survey and observation method and the author's practical experience in the duration from 6/2020 to 6/2021 through 169 companies, including 107 foreign capital invested companies and 62 Vietnamese-owned companies in Ho Chi Minh City and other provinces such as Dong Nai, Ba Ria - Vung Tau, Binh Duong (Table 1).

Table 1: The quantity of company being observed

Company located at	Foreign invested company		Vietnamese owned company		Total	
	Qty	%	Qty	%	Qty	%
HCMC	9	8.4%	39	62.9%	48	28.4%
Binh Duong	35	32.7%	2	3.2%	37	21.9%
Dong Nai	27	25.2%	3	4.8%	30	17.8%
Ba Ria - Vung Tau	14	13.1%	5	8.1%	19	11.2%
Other	22	20.6%	13	21.0%	35	20.7%
Total	107	100%	62	100%	169	100%

The following difficulties and problems faced by the companies establishing and applying the ISO 9001:2015 are the results from the author's survey and observation with practical experience. The research was done in the duration from 6/2020 to 6/2021 through 169 companies in Vietnam, including 107 foreign capital invested companies and 62 Vietnamese - owned companies (Table 1). The detailed results of research are mentioned in the Table 2.

The first problem is relating to "purpose of application". The application of a quality management system is a strategic decision that can help an organization to improve its overall performance with sustainable development (ISO 9001:2015). The purpose of this standard is to assist an organization in establishing, implementing, maintaining and improving a quality management system to provide consistent quality products and services that comply with customer and regulatory requirements, which brings the long-term benefits to the businesses. However, the enterprises have not set a long-term goal of applying this standard as a management tool to stabilize and improve their operational processes. They only consider the application for ISO 9001:2015 certification or other certifications as a temporary goal to meet the customers' requirements for signing purchasing contracts. In other cases, the enterprises obtain the certificates only to deal with the procedural requirements in competitive contracts such as participating in bidding for works or projects; or just to advertise and embellish for the business. Therefore, they do not invest efforts and costs for setting up and applying the quality system complying with the orientation of standard. Besides, when the target customers or the short-term goals are no longer available, the enterprises are not interested in applying, maintaining and improving the system anymore.

There are 78% of foreign invested companies and 84% Vietnamese owned companies getting this problem (Table 2). To get out of the problem, the development and application of the quality management system should be considered as long-term strategic goal and as a management tool to stabilize and improve the company's operational processes.

The second difficulty is "Time for setting up the quality management system". To establish a relatively complete quality management system and operate smoothly, businesses need to invest time of average about 6 months to 1 year or more and depending on whether the business hires a consultant or not. However, the companies or enterprises rush to set up the system in order to get ISO 9001:2015 certification urgently because they are going to sign a contract with customers; or the deadline for submitting bids is nearly ending; or an advertising campaign is coming up, etc. Therefore, enterprises do not invest enough time to review the completeness of the quality system and its operation effectiveness. They just set up a basic quality management system that the most fundamentally meets the standard requirements for being certified.

There are 73% of foreign invested companies and 77%

Vietnamese owned companies getting this problem (Table 2). The solution suggested is that the reasonable duration of time should be invested for setting up the quality management system, normally at least 6 months. The completeness of the system should be checked before being put under implementation and its operation effectiveness reviewed for corrective action timely.

Table 2: The Companies have Difficulties and problems in establishing and applying ISO 9001:2015

No.	Difficulties and problems	Foreign invested company (107 Companies)		Vietnamese owned company (62 companies)	
		Qty	%	Qty	%
1	Purpose of application	83	78%	52	84%
2	Time for setting up the Quality Management system	78	73%	48	77%
3	Documented information system establishment	62	58%	21	34%
4	Leadership and commitment	52	49%	38	61%
5	Awareness and engagement of people	45	42%	42	68%
6	Resources and training activities	37	35%	47	76%
7	Application	37	35%	43	69%
8	Taking corrective actions	32	30%	26	42%
9	Monitoring, measuring and evaluation	28	26%	47	76%

The most important problem is related to "Leadership and commitment". One of the requirements when developing and applying a quality management system is "Leadership and commitment". The article 5.1 of the ISO 9001:2015 standard says that the top management shall demonstrate leadership and commitment by engaging, directing and supporting persons to contribute to the effectiveness of the quality management system. However, the management has not paid much attention to the importance and effectiveness the quality management standard application as well as because of the above-mentioned temporary goals, the management has not really shown the commitment yet. For examples, they are including but not limited to the bellows.

The management has not recognized yet the implementation of the quality management system according to ISO 9001:2015 as a management tool. Enterprises consider this as an individual and independent system rather than a part of the company's overall management system.

The management has not paid proper attention yet to the long-term benefits of applying ISO 9001:2015 in stabilizing production processes in particular and other operating processes to help businesses reducing risks, non-quality costs and shadow costs incurred in business.

Management has not focused on the importance of investing costs in preventing defects yet. The costs invested for education and training are the most reasonable and importantly preventive costs. Error prevention is the core issue to produce consistent quality.

Last but not least, not all of the management members fully and clearly understand the standard requirements that they are applying, even misunderstanding the contents and target application.

There are 49% of foreign invested companies and 61% Vietnamese owned companies getting this problem (Table 2). The solution for this issue is that the management should engage and support persons in developing, implementing and contributing to the effectiveness of the quality management system to plan and organize the necessary duties properly for participants.

The next difficulty is the "Awareness and engagement of people". ISO 9000:2015, the standard on Quality Management Systems - Fundamental and Vocabulary (fourth edition), has mentioned 7 quality management principles and "engagement of people" is one of them. The same as the management, the employees have not fully and clearly understand the standard requirements, too. They have not recognized the importance and the benefits affecting the effectiveness of the processes which they themselves are operating or participating in when applying the quality management system. They usually perform works as routine and experience with being afraid of change, so they feel uncomfortable when being asked to follow newly issued working procedures. They often consider the application of standard requirements as increasing more tasks which are not their business, so they often avoid participating in or ignoring doing additional tasks. They have not yet realized that the working procedures helps them to perform and manage their work more consistently and systematically, as well as solving problems more thoroughly and effectively.

There are 42% of foreign invested companies and 68% Vietnamese owned companies getting this problem (Table 2). To solve this problem is to focus on training and guiding people how to do. The awareness of the standard requirements and operation should be done for all levels from management to workers. The extent of training content should be provided depending on each level. All of people should encourage to participation in setting and implementing the system to understand the processes and their relationship.

Another difficulty is providing “Resources and training activities”. As mentioned above, the establishment and application of a quality management system should follow quality management principles such as "Leadership" and "engagement of people" (ISO 9000:2015). However, the management often entrust this work to one employee or a small group of employees. These employees and a number of departmental representatives have participated in awareness and internal auditor training. When setting up workflows, only this group of employees many times has to do the tasks on behalf of other departments, even though they sometimes can't fully understand the processes being set up. Or conversely, the trained persons have not participated in preparing the processes but leaving the untrained people to do that. In addition, the training time is shortened because the enterprises rush to apply for certification, so employees are not trained well enough to understand the standard contents or how to set up and apply the system effectively. Another problem is that the organizations have not invested in training or organized regularly internal training.

There are 35% of foreign invested companies and 76% Vietnamese owned companies getting this problem (Table 2). To get out and improve the system, the training should be invested enough, and training cost should be considered as costs of preventing defects. Resources should be provided timely for operation of the system to get better performance.

Another difficulty is “Documented information system establishment”. According to the article 7.5.1 of ISO 9001:2015, the extent of documented information system can differ in organizations due to the organization’s size and type of activities, processes, products and services; the complexity of processes and their interactions; and the competence of persons. However, the organizations have still created too many unnecessary documents and the document control becomes complicated and cumbersome. The documented information system set up in the organizations still has the problems, including but not limited to the followings.

Confusing and unclear translation: this problem is often encountered in group of companies or organizations having mother in foreign countries that have already developed and applied standards. The documented information system written in mother language is transferred to the organization located in Vietnam and is translated into Vietnamese. However, the translators who know the language do not understand the terminology and technical terms written in the documents; and the experts who understand the terminology do not know the language, so the translation of the documents is often inaccurate and hard to comprehend.

Documents inconsistent with operation: besides the incorrect and confusing translation, some other documents of the parent company with the way of operation are not suitable for the actual conditions of enterprises operating in Vietnam (because the parent company has existed and developed for a long time). They are not allowed to reduce or modify by the parent company. This is one of the reasons for the discrepancy between what is mentioned in the documents and the actual operation in the organization. The existence of non-applicable documents also contributes to the complexity of the documented information system. Besides, from the first stage of building a document system, the domestic enterprises have entrusted to consulting organizations. The consultant often takes the available sample set of documents from other companies, amending a little bit and then gives them to the client company to apply. These documents have not been checked carefully by enterprises before issued, so there may be inconsistencies between documents and actual operations.

Documents uncombined: Many businesses have been applying a combination of many management systems according to international standards such as ISO 9001:2015, ISO 14001:2015, ISO 45001:2018, etc. These standards have a similar structure to make it easy to be applied. The organizations may develop and apply these standards at different times. However, they do not take time to review and combine similar documents to achieve a simple and consistent documented information system applied for all systems.

There are 58% of foreign invested companies and 34% Vietnamese owned companies getting this problem (Table 2). A better solution suggested that the documents should be simple, concise but including enough contents and easy to understand for effective application.

“Application” after establishment the quality system is another difficulty. Due to the lack of reasonable investment in time, training and resources for the system setup phase, the organizations have faced, including but not limited to the following problems when putting the system into operation.

Employees still work as habit and don't care what's written in the documents. That’s why the organization’s documentation system has never been read although it has been certified to ISO 9001: 2015 for a long time. Therefore, the employees do not find out that they are doing things differently from regulations until the assessment body raises out.

In addition, if employees are required to follow the rules to meet the requirements of the system, they are unwilling or do it in a sketchy way. They are also not interested in reviewing documents carefully during operating process to have timely adjustments and improve the operation simpler and more effective. It is the employees who get benefits from these results.

There are 35% of foreign invested companies and 69% Vietnamese owned companies getting this problem (Table 2). For better performance, the documents should be regularly checked and reviewed for their effectiveness and timely revised to avoid the difference between the regulation and operation. The monitoring and measuring activities should be done frequently to check if any non-conformity happens.

“Taking corrective action” is another problem. During the application and operation of the system, if there is any nonconformity determined and needed corrective action, the root causes will be analyzed and corrective actions proposed to be taken. However, the problems are not always properly analyzed then appropriate corrective actions can not be taken to eliminate the root causes of nonconformity, which cause its recurrence.

There are 30% of foreign invested companies and 42% Vietnamese owned companies getting this problem (Table 2). Appropriate corrective actions should eliminate the root causes of nonconformity to prevent recurrence of non-conformity is one of solution suggested for this problem.

The last, but not least, difficulty found in this survey is relating to “Monitoring, measuring and evaluation activities”. In the process of applying systems, it is indispensable to monitor, measure and evaluate the performance. An internal auditor team conducts a planned internal audit to determine the effectiveness of the system. However, the auditors have not been trained well in skills and do not have enough experience in auditing so the audit results are not reliable enough for suitable improvement actions.

There are 26% of foreign invested companies and 76% Vietnamese owned companies getting this problem (Table 2). To improve the performance of the quality management system, the internal audit is one of the monitoring and evaluating activities to provide information on whether the system conforms to the requirements of the organization itself or of this International Standard; and on the effectively implemented and maintained performance (ISO 9001:2015). Auditor team play an important role in this activity. Therefore, the internal auditors should be fully trained with audit skills and technique with enough required qualification. The evaluation for auditors’ competences should be done regularly to ensure that they are complying with company and work requirements.

Despite the limitations of 9 proposed solutions, the organizations’ quality management systems are basically compliant with the requirements of the ISO 9001:2015 standard. Then, they can apply the certificates of conforming to the ISO 9001:2015 for their organizations. The non-conformity found during the certification audit by the Third party (Certification Body) will be taken corrective actions for improvement and will be checked by the Certification Body in the surveillance audit every year in the 3-year circle of certification.

4. Conclusion

The article contents are basing on the results of the survey and observation on the sample of 169 companies located in Vietnam which have been setting up and applying ISO 9001:2015. In this article, the authors have worked out 9 difficulties and problems with that the companies have to deal. The results shows that most of the companies have difficulty with “Purpose of application” including foreign capital invested companies and Vietnamese-owned companies. The second rank is the problem relating to “Time for setting up the Quality Management system” for both types of companies. The lowest rank is problem of "Monitoring, measuring and evaluation" for foreign capital invested companies and "Documented information system establishment" for Vietnamese-owned companies. The solutions for these difficulties and problems have been suggested by the authors, too. The above mentioned solutions are the very basic ones to assist the companies to approach the orientation of the ISO 9001:2015 standard in improving the system and providing products and services with consistent quality according to customers' requirements and relating regulatory and statutory requirements.

With the above results, this article is basing on the survey result of a small sample of 169 enterprises including foreign invested and Vietnamese ones. This sample size has not presented for all enterprises in various main industries yet. The proposal for a larger sample size of enterprises presenting for different industries is the orientation for the next research with the desire to achieve more specific results.

References

Beattie, K. R., & Sohal, A. S. (1999) Implementing ISO 9000: A study of its benefits among Australian organizations. *Total Quality Management,* 10, 95-106.

Boiral, O. (2012) ISO 9000 and organizational effectiveness: A systematic review. *The Quality Management Journal*, 19, 16-37.

Casadesus, M., & Gimenez, G. (2000) The benefits of the implementation of the ISO 9000 standard: Empirical research in 288 Spanish companies. *The TQM Magazine*, 12, 432-441.

Casadesus, M., & Karapetrovic, S. (2005) Has ISO 9000 lost some of its luster? A longitudinal impact study. *International Journal of Operations & Production Management,* 25, 580-596.

Chow-Chua, C., Goh, M., & Wan, T. B. (2003) Does ISO 9000 certification improve business performance? *International Journal of Quality & Reliability Management,* 20, 936-953.

Corbett, C. J., Montes-Sancho, M. J., & Kirsch, D. A. (2005) The financial impact of ISO 9000 certification in US: An empirical analysis. *Management Science*, 51, 1046-1059.

Department Of Science And Technology - Ministry Of Industry And Trade (2020) *Survey on the number of ISO certificates issued in Vietnam by industries* (23/11/2020).

Douglas, A., Kirk, D., Brennan, C., & Ingram, A. (1999) Maximizing the benefits of ISO 9000 implementation. *Total Quality Management*, 10, 507-513.

Dunu, E. S., & Ayokanmbi, M. F. (2008) The impact of ISO 9000 on the Financial Performance of Organizations. *Journal of Global Business Issues,* 2, 135-144.

Han, S. B., Chen, S. K., & Ebrahimpour, M. (2007) The impact of ISO 9000 on TQM and business performance. *Journal of Business and Economic Studies,* 13.

Heras, I., Dick, G. P. M., & Casadesus, M. (2002) ISO 9000 registration's impact on sales and profitability: A longitudinal analysis of performance before and after accreditation. *International Journal of Quality & Reliability Management*, 19, 774-791.

International Organization for Standardization (2015) ISO 9001:2015 - Quality Management Systems - *Requirements*

International Organization for Standardization (2015) ISO 9000:2015 - Quality Management Systems - *Fundamental and Vocabulary*

Martinez-Costa, M., & Martinez-Lorente, A. R. (2007) A triple analysis of ISO 9000 effects on company performance. *International Journal of Productivity and Performance Management*, 56, 484-499.

Sharma, D. S. (2005) The association between ISO 9000 certification and financial performance. *The International Journal of Accounting,* 40, 151-172.

Starke, F., Eunni, R. V., Fouto, N. M. M. D., & Angelo, C. F. (2012) Impact of ISO 9000 certification on firm performance: Evidence from Brazil. *Management Research Review*, 35, 974-997.

Vault IQ (2019), ISO Spotlight: *What do these 10 standards mean for your business?* (10 Dec. 2019).

Withers, B. E., & Ebrahimpour, M. (2000) Does ISO 9000 certification affect the dimensions of quality used for competitive advantage? *European Management Journal*, 18, 431-443.

APPLYING LEAN IN A GARMENT MANUFACTURING COMPANY: A CASE STUDY

Nguyen Huu Phuc[1] and Trinh Thi Thuy An[2]

[1] Nguyen Huu Phuc, Ho Chi Minh City University of Technology , nhphucise@hcmut.edu.vn

[2] Trinh Thi Thuy An, Ho Chi Minh City University of Technology, an.trinh99@hcmut.edu.vn

Abstract. The research discusses how the Lean manufacturing could be adopted in textile industry using a sewing plant case study. The purpose of this research is indentifying wastes of operations in sewing line in oder to eliminate non-value added activities and minimize idle time between workstations, may also reduce number of operators in sewing line. Value Stream Mapping (VSM) tool is firstly using to address various wastes in the present process. To eliminate non-value added activities, the working design, line balancing, kanban system and FIFO system tool are implemented. The outcome of this research is reducing 46.6% in lead time and increasing 21.08% in line balancing efficiency in addition to 69.5% reduction in work in process time. So that, the proposal of improvement the process production of Polo shirts allow guarantee the productivity and the competitiveness leading to increase order of customer.

Keywords: Lean manufacturing; VSM; Work design; Line balancing, Kanban, FIFO.

1. Introduction

Elite Long Thanh Co., Ltd of TMI group specializing in manufacturing of Elite Long Thanh process of sportwear for the ADIDAS brand as t-shirt, jacket, pants,…

Currently, the manufacturer is facing with the essential problem that is the poor productivity of sewing line as unexpected, thereby leading to reduce the order fulfilment rate as well as the competitive advantage compared with other companies in the same field. So, with the purpose of increasing productivity and get more orders from customer, Lean manufacturing is the suitable method to eliminate wastes, reduce lead time in the process.

The manufacturer is processing many different product lines on orders of customer, in which Polo shirts is the main line that has the highest production rate and is stable in all seasons. There are many product families with the same process but only a few different components. PWS product family is chosen with the highest order. So, TM1923F21 has the basic production process that is as a research object for the improvement project.

2. The current state value stream map

VSM is a lean manufacturing's visualisation tool which can aid this task by showing the flow material along with information as the product passes through the transformation process. Consequently, the current state VSM is used to fully identify various wastes of the current process; and then find out the method to eliminate it.

2.1. Data collection and drawing the current state map

Information flow includes information that tells each process what to produce and when to produce it. In this case, specific information that included requirements of raw material, suppliers' leadtime and shipping schedules were collected. In addition, customer demand data as daily production demand was captured. On the other hand, material flow provides various essential characteristics of the process to drawn VSM that included cycle time of each station, work-in-process before station, number of operatiors involved of each station and machine.

Specifically, cycle time of each station is collected by work sampling that use stopwatch and camcorder to calculate the data. The manufacturing facility operated on a six-day working week and employed 1 shift of 8 hours 45 minutes each, with a break of 45 minutes, thereby the avaible production time (APT) is 8 hours. Daily production demand was predicted to be 650pcs

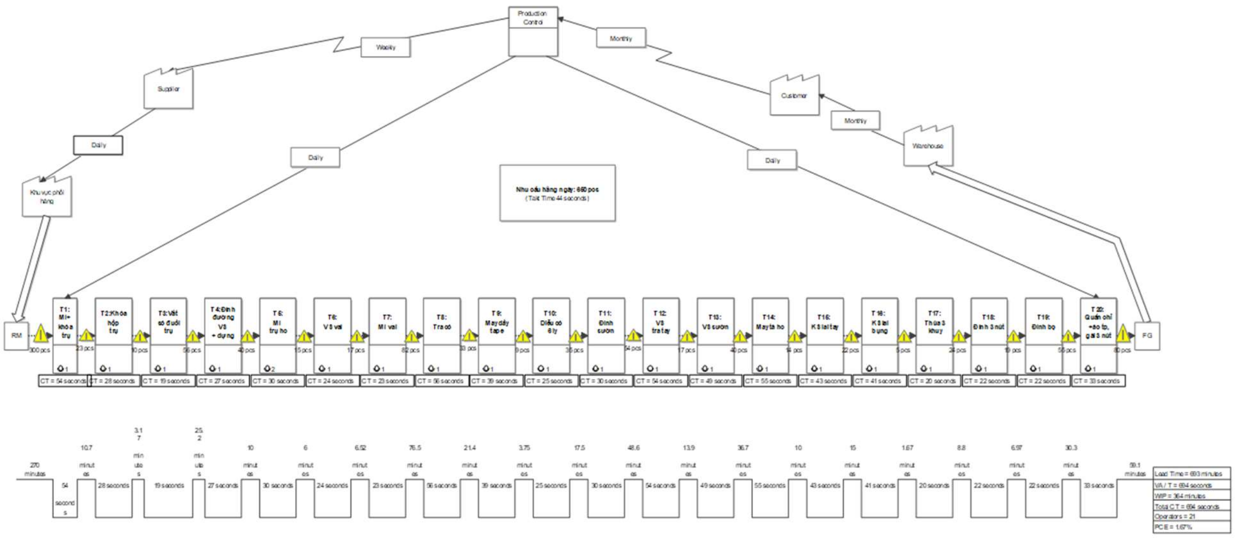


Figure 1. The current state VSM.

2.2. Analysis of current state VSM

VSM parameters are calculated and the VSM illustration is drawn by IGrafx (see *Figure 1* illustrates The current state VSM and Paremeters of the current process are summarized in *Table 1*).

Table 1. The paremeters of the process

Lead time	693 m
VA/T	11.56 m (694 s)
WIP	364 m
PCE	1.67 %
CT	56 s
Takt time	44
Operators	21
Stations	20

According to *Figure 1*, it is evident that a push system that the upper station processes semi-finished products and passes them to the next station without regard to the inventory of semi-finished products in line, so lead to high amount of WIP .

A lots of station as station have very higher than the takt time and hence these operations required improvements or capacity addition to meet demand. (see *Figure 2*).

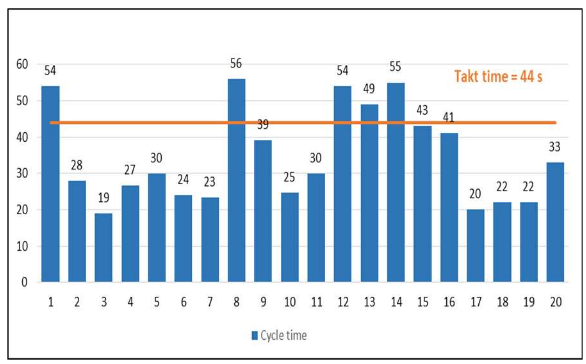


Figure 2. The CT in relation to the takt time

Accordingly, the following tools are adopted :

- Using work design, line balancing tool is to decrease the idle time, thus solve the CT higher than the takt time problem.
- Then using Kanban system and FIFO system is to decrease WIP in the production process to deal with the long lead time problem.

3. Work design and line balancing

3.1. Work design

Work design is applied by observe, studying and analyzing the basic movements of operators of each station as sewing operations, taking semi-finished products during work execution to optimize and eliminate the movement wastes, thus reducing the processing time.

Based on *Figure 2* , priority is given to improve the operation of stations with cycle time higher than takt time including station 1, station 8, station 12, station 13, station 14.

By the way, the cycle time is significantly reduced. As a result, the Cycle time of all stations met Takt time (see *Figure 3*).

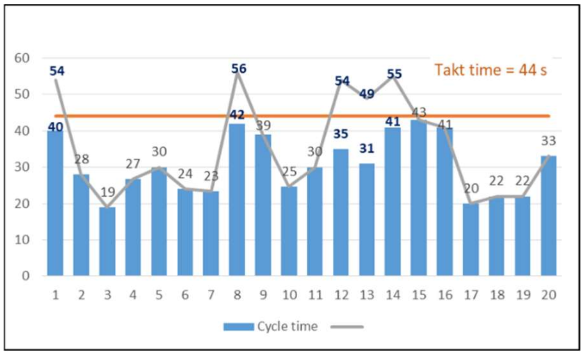


Figure 3. The Cycle time after applying work design

3.2. Line balancing

However, the processing time diffirence between stations is quite high. Therefore, continue to use the method of line balancing to balance the cycle time of stations leading to reduce the idle time and synchronize the division of labor. Minimum number of stations is calculated according to formula (1) as:

$$N_m = \frac{Total\ Cycle\ time}{Takt\ time} \qquad (1)$$

With total cycle time is 607s and takt time is 44s, it can be seen that the optimal number of stations is 14 stations. But due to the limited knowledge of the textile industry, it is only possible to solve the agglomeration of stations in the best possible way. Split and merge stations as follows:

- Combine station 3 with station 6 into one station called Station 2 with 1 worker performing 2 stages of "Oversew the placket and the shoulder".
- Separation of station 4 includes 2 stages, then combine 1 stage with station 2 called station 3 with 1 worker performing 2 stages. Similarly, combine 1 stage with station 5 called station 4 with 2 worker performing 2 stages.
- Combine station 17 and station 18 called station 15 with 1 worker.

As a result, the total number of stations is only 17 and the total number of workers is only 18 (see *Figure 4)*

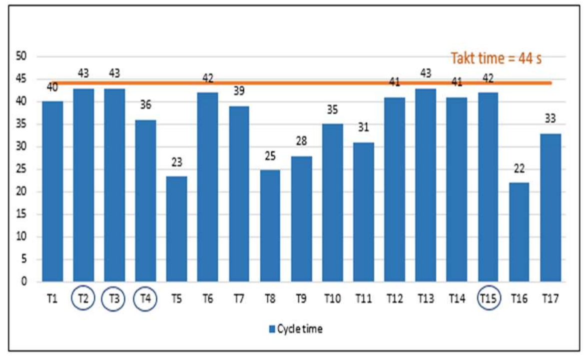


Figure 4. The Cycle time after line balancing

4. Kanban system and FIFO system

4.1. Kanban

The current 'push' system causes a large amount of inventory in the sewing line at some stations that has has a larger cycle time than the previous station.

Implement the new Kanban system is to control WIP well with the stated goal of reducing the total amount of inventory between stations and reducing the lead time.

In this case, as observed, the layout of the space between the stations on the line is located close to each other and the travel distance is short, so the single-card Kanban system is applying. Kanban is placed between two stations when the former has more production capacity than the latter (see *Figure 5*).

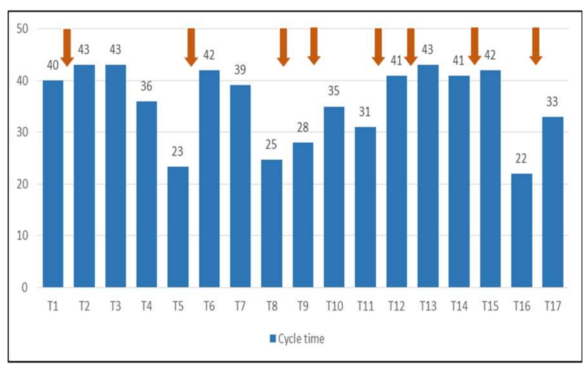


Figure 5. Kanban-card location

The number of Kanban is calculation according to the formula (2) as :

$$N = \frac{DL\,(1+\alpha)}{Q} \qquad (2)$$

Where,

- N : Number of kanban
- D: Daily demand (650 pcs/day)
- L: Leadtime
- α: Safety stock (α =0.1)

- Q: Kanban size (Q=10)

Because WIP is transferred between stations by the leader, so Lead time is equal to the Cycle time. Accordingly, the number of Kanban is caculated (see *Table 2*).

Table 2. The number of Kanban

Station	N
T1 & T2	1
T5 &T6	1
T8 &T9	1
T9 & T10	1
T11& T12	1
T12 & T13	2
T14 & T15	2
T16 & T17	1

4.2. FIFO system

After using the Kanban system, continue to design the FIFO system, which is used to prevent diffence of Cycle time of the stations. In this case, there is still the inventory of WIP from low capacity station to high capacity stations.

Set FIFO between stations with Cycle time of previous station greater than cycle time of following station (see *Figure 6).*

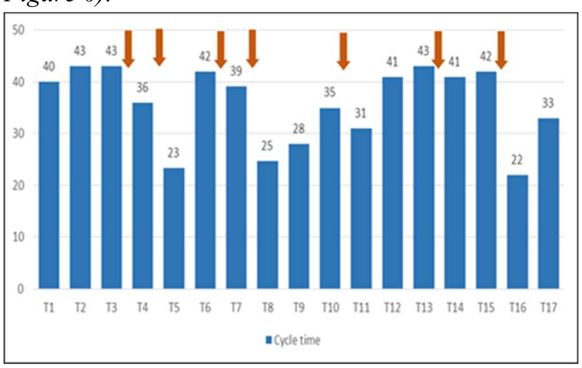


Figure 6. FIFO location

Use inventory sizing software provided by Christoph Roser to calculate the size of the FIFO lane. *Table 3* show the size FIFO.

Table 3. The size FIFO

Station	FIFO
T3&T4	7
T4&T5	9
T6 &T7	8
T7&T8	8
T10& T11	9
T13& T14	7
T15 & T16	8

5. The future state VSM

After apply the improve project, drawing the future state VSM to evaluate the improved process as well as analyzing the impacts of appling Lean on production. (see *Figure 7* illustrates The future state VSM and *Table 4* show the parameters of the future state VSM).

Table 4. The parameters of the future state VSM.

Lead time	370 m
VA/T	10.11 m (607 s)
WIP	111 m
PCE	2.74 %
CT	43 s
Operators	18
Stations	17

6. Conclusion

With the application of Lean, these tools (VSM, Work design, Kanban, FIFO) were used, achieving the significant improvement of the productive process, indentifying wastes of operations in sewing line in oder to eliminate non-value added activities and minimize idle time between workstations, may also reduce number of operators in sewing line. Accordingly, by implementing of this proposal, a total production time of 607 second is obtained; no station has fully cycle time greater takt time that the cycle time reduce from 56(s) to 43(s) and satisfied the takt time as 44(s); PCE ratio is improved 1.07%, increasing form 1.67% up 2.74%; lead time approximately decreases by 46.6% in addition to WIP time significantly decreases by 69.5%.

So that, the proposal of improvement the process production of Polo shirts allow guarantee the productivity and the competitiveness leading to increase order of customer.

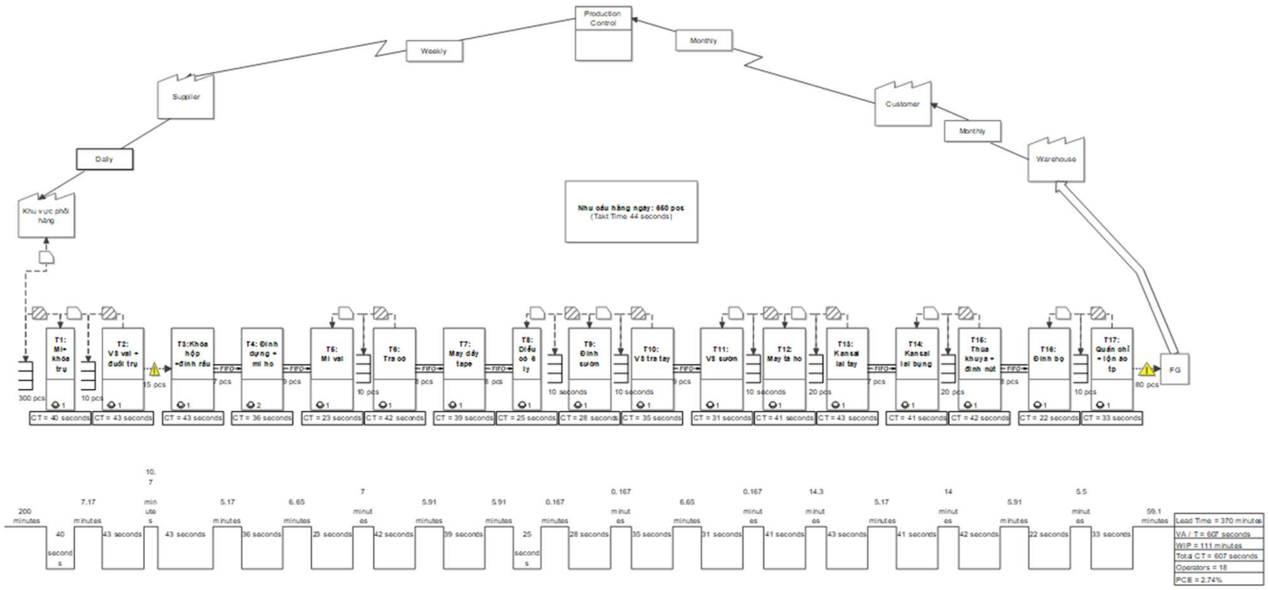


Figure 7. The future state VSM.

Reference

Nguyễn Như Phong, *Sản xuất tinh gọn*. National University of HCMC Publisher, 2016.

Hồ Thị Minh Hương, Lê Song Thanh Quỳnh, *Thiết kế chuyền may công nghiệp.* National University of HCMC Publisher, 2013.

Kazi Mohammed Saidul Huq, Konstantinos Mitrogogos. *Impact of Lean manufacturing on process industries.* Master of Business Administration Program, 2018.

Nguyễn Như Phong, Nguyễn Hữu Phúc, Võ Văn Thanh, Hà Thị Thúy Vân. *Applied value stream mapping lean production system Clipsal VietNam Co.* Science & Technology development, Vol 18, No.K1- 2015.

Álvaro José Almeida Borges Pereira Pinto, *Lean management: Wood Manufacturing Industry Case Study*, Master in Industrial Engineering and Management.

Domingo, R., Alvarez, R., Melodía Peña, M. and Calvo, R. (2007). *Materials flow improvement in a lean assembly line: a case study*. Assembly Automation, Vol. 27 No. 2, pp. 141-147.

A STUDY OF INLAND WATERWAYS LOGISTICS IN SOUTHERN VIETNAM

Vo Trong Cang[1,2], Cao Viet Hieu[1] and Nguyen Anh Tuan[1,3]

[1] *Binh Duong University, Binh Duong, Vietnam*

[2] *IMALOG, Binh Duong University, Binh Duong, Vietnam, vtcang@bdu.edu.vn*

[3]*TYVY Yacht and Education Co., Ltd., tyvy.net@gmail.com, natuan87@bdu.edu.vn*

Abstract: The paper concentrates on analysing the general situation in logistics and supply chain management **(SCM)** in the South of Vietnam, especially in the business connection between Binh Duong province and Ho Chi Minh city besides other surrounding regions. In addition, it is crucial that the decisive factors including inland waterways transportation infrastructure, logistics policies, the economic environment, and professional workforce in logistics and SCM will be illustrated and considered seriously. As a consequence, it is of utmost importance that the workable solution will be proposed based on both the advantages and disadvantages of the economic and social environment in southern Vietnam.

Keywords: Logistics, Supply chain management, Southern Vietnam, Inland waterways transportation.

1. Introduction

In Vietnam, there is a sophisticated system of inland waterways that play an essential role in the economical structure, especially in freight transport. According to the report conducted by the World Bank in 2019, it constitutes 20 per cent of the total commodities transported within Vietnam in comparison with 5% to 7% of the shipments delivered by the means of inland waterways in China, the United States of America and European Union (Hoang Anh Dung, 2019). Besides that, Vietnam possesses the natural advantage of inland waterways systems with 19,000 km in association with about 3,200 km of coastal waters, and it has not been operated and underestimated (Yin Lam, 2019).

Regarding the environmental pollution due to the activities relevant to logistics and supply chain in goods transport, it is unarguable that the inland waterways and coastal transportation contribute to a smaller percentage of greenhouse gases than other transportation methods (Hoang Anh Dung, 2019).

In terms of the cost of goods transportation, the enormous volume of merchandise, especially super heavy, over-dimensional and unusual freight, can be transported safely and efficiently with reasonable prices by a large quantity of cargo barges on inland waterways, particularly the involvement of many local private enterprises. As the result, the local government in Binh Duong province considers the inland waterways system as a principal role in economic development. Tan Cang - Long Binh ICD JSC has been established since 2007 within total areas of 230 ha, there is much more incentive for local industrial areas in Dong Nai province, Tay Ninh province, and Binh Duong province to connect with the logistics systems in Ho Chi Minh city and particularly Ba Ria-Vung Tau province which has Cai Mep – Thi Vai port cluster (JSC, n.d.). The capacity of Cai Mep – Thi Vai can load the import and export products from the ultra-large cargos with a deadweight of more than 214,000 tons, and the shipments of containers from Ho Chi Minh city to Cai Mep – Thi Vai port which account for 90 per cent were transported by inland waterways vessels (Luis C. Blancas, 2014). As a consequence, it opens the main gate for a remarkable range of local products to enter the international marketplace through inland waterways systems.

Nowadays, the municipal authorities in Ho Chi Minh city and generally in other cities in Vietnam have drawled considerable attention to the role of logistics and SCM in not only the reinforcement of the competitive strength and the development of the local economy, but the improvement of the capacity in goods transportation and the optimization of the economic and social resources as well. Take Binh Duong province as an excellent example, the local administration emphasised that the substantial reduction of transportation cost by utilising the integrated logistics process is underlying fundamentals in enhancing the competitive strength of the enterprises based on the improvement of the inland waterways transportation infrastructure (Department of Information and Communication of Binh Duong, 2018).

2. Information technology integrated into logistics and supply chain management

In order to fulfil customers demand on time, it is indispensable that figure 1 demonstrates that the logistics management process has to manage effectively and resourcefully an enormous amount of critical information ranging from the supplies to the distribution of the products (Christopher, 2011). Recently, there have been many contemporary concepts of logistics that deliberate on the complete integration of all aspects in business operation in one solution. To illustrate the trend of customer behaviour and optimise the inventory management, marketing and manufacturing procedures are no longer considered separately compared with the conventional methods of the organization administration in the past (Christopher, 2011).

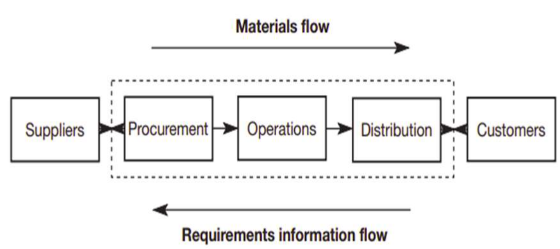


Figure 11. Logistics management process (Christopher, 2011)

There is an overwhelming number of definitions of supply chain management but all of them which have three main groups are composed of the upstream suppliers, the firm, and the downstream parties as in figure 2 (Stanley E. Fawcett, 2014). Based on the extension of the logistics companies, the model of SCM will be more complicated, and hence, the managers should implement properly the workable systems of the logistics and SCM.

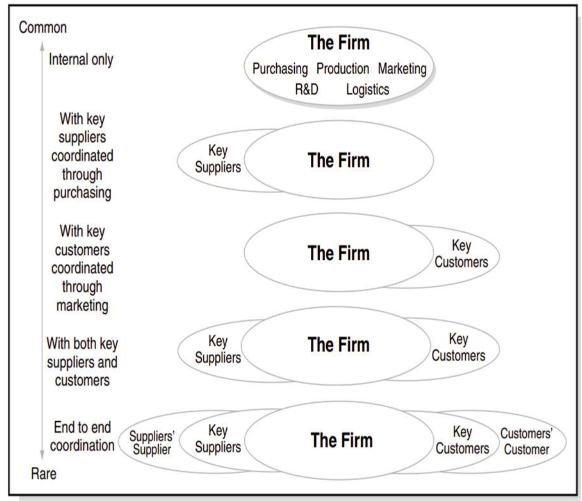


Figure 12. Models of supply chain integration (Stanley E. Fawcett, 2014)

Since the accelerated development of information technology including big data, blockchain technology, and the internet of things (IoT) has occurred recently, the realistic prospects for the integrated logistics and supply chain system can be considered for the revolutionary transformation of the conventional logistics organisations. Concerning other cutting-edge technologies in computer-aided design, computer-aid manufacturing, product lifecycle management, they also contribute to the dramatic transformation in logistics and SCM.

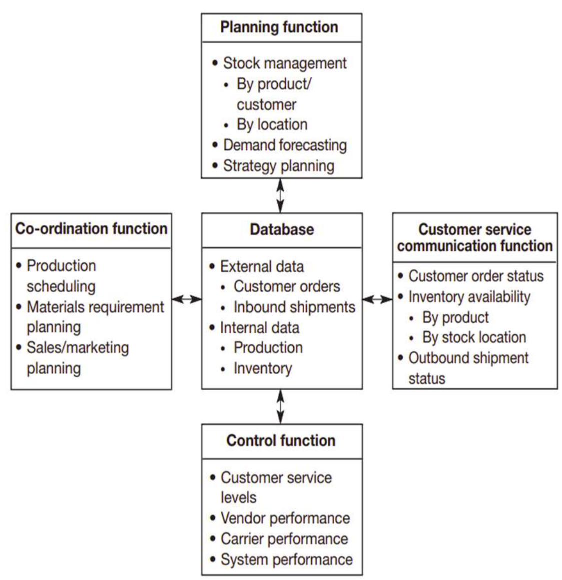


Figure 13. Application of information technology in logistics (Christopher, 2011)

It has become apparent that the internet, especially the 5th generation of the wireless network, will enable the logistics business to integrate much more data into their logistics platforms so that all business partners in the logistics systems can cooperate effectively, and hence reduce significantly the cost of logistics by optimizing the schedule of the replenishment and estimating precisely customer's forthcoming demand in real-time. As can be seen from figure 3, it is essential that the core of IT systems applied in logistics and SCM is always the online database which contains not only customer's requirements and inbound shipments but also the data of the production and the inventory management. The data centre will receive the updated information simultaneously with 4 critical functions including planning, customer service communication, coordination and control (Christopher, 2011).

3. Implementation of integrated logistics and SCM solutions in inland waterways

As discussed in the introduction, the differential advantage of the inland waterways systems with in 19,000km in Vietnam should be considered as the crucial factor in developing the modern integrated inland transportation system including road transport, inland waterways shipping, and air cargo shipments. It can be illustrated from figure 4, there would be many benefits that Vietnam has to integrate the inland waterways shipment system into inbound and outbound logistics solutions, especially the transportation in raw materials and the distribution to warehouse and wholesalers, causing to diminish substantially the cost of the merchandise transportation (Paul R. Murphy & Knemeyer, 2018).

Besides that, not only are the warehousing, the inland ports, seaports, and other related facilities composed of the integrated logistics systems, but resources management, production and operation, packaging and assembly service, and business administration should get intimately involved as the essential factors of logistics and SCM (Paul R. Murphy & Knemeyer, 2018). The proposal of the logistics systems for the inland waterways shipping was suggested for Egypt in 2006, it will comprise two essential groups which are unmanageable (minus sign " – ") and manageable (plus sign "+") (Islam El-Nakib, 2006).

It can be clearly seen from figure 5 that there is an enormous quantity of related factors in inland waterways logistics systems that should be considered and assessed properly. Some of them will be unstable because of the fact that not only have the advanced technologies been utilised so far, but the covid-19 pandemic influences the demand and supply in the global and local marketplace by the unpredictable ways, and hence these will bring the challenge and opportunities for the innovation in logistics and SCM systems.

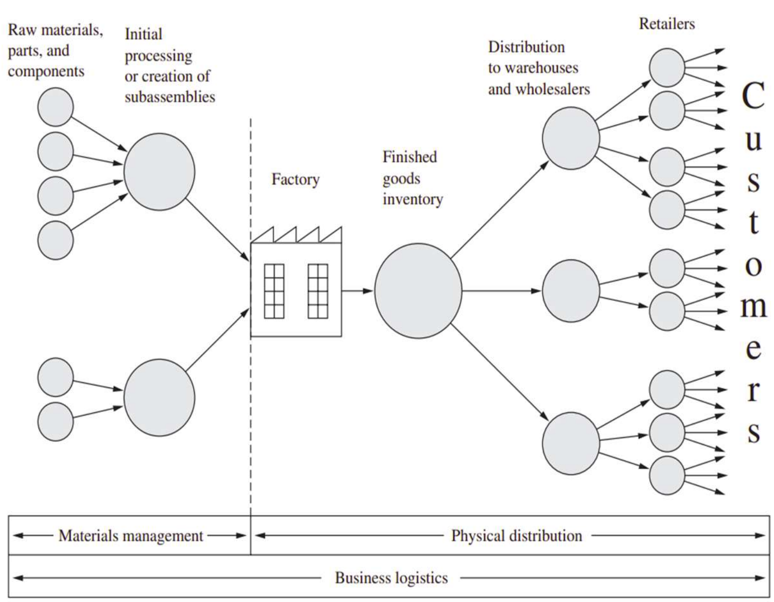


Figure 14. The management process of inbound and outbound logistics solutions (Paul R. Murphy & Knemeyer, 2018)

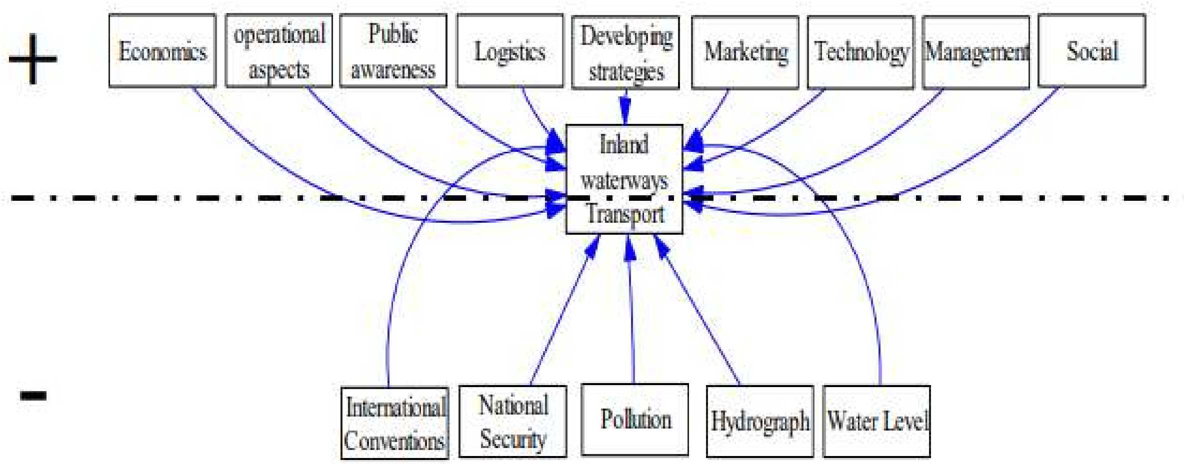


Figure 15. The innovative concept of the inland waterways logistics systems (Islam El-Nakib, 2006)

As is illustrated in figure 5, the proposed concept of the integrated logistics solution for the inland waterways systems will be directly influenced by the economic and social conditions, the mutual effects of the government's legal framework and private enterprises' logistics activities, the geographical features, and other relevant issues. Particularly, the development of the waterborne transportation system depends on both the vision of centralised government and regional authorities in association with the international conventions, and it will contribute to lessening the enormous amount of environmental pollution by reducing the road traffic congestion, and decreasing the rate of noise and air pollution in relation with the truck industry (Paul R. Murphy & Knemeyer, 2018).

Intermodal and multimodal logistics solution

It is crucial that the contemporary logistics systems will be composed of intermodal and multimodal transportation. Due to the containerization in recent decades, as represented in figure 6, the intermodal application plays a dramatic role in standardizing the size of freight containers in the twenty-foot equivalent unit (TEU), and as a result, it offers the economic advantage by its convenience in handling simultaneously the standardized containers amongst ocean terminals, deep seaports, railway systems, road transport, and inland waterways (Paul R. Murphy & Knemeyer, 2018).

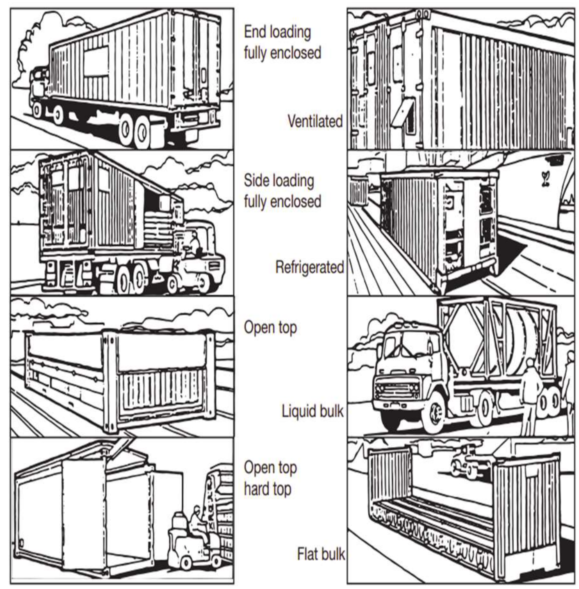


Figure 16. A wide variety of standardised containers in intermodal transportation (Paul R. Murphy & Knemeyer, 2018)

Figure 17. Centerpoint Intermodal Center (Luis C. Blancas, 2014)

An outstanding example of the implementation of intermodal logistics is that a successful enterprise is known as CenterPoint Intermodal Center (in figure 7), which was located in Ellwood (about 40 miles to the Southwest of Chicago), Illinois, the United States, and it was established by the active involvement of both governmental authorities and private businesses in the early stages of the ambitious project in 2002. In addition, by implementing the multimodal logistics system (figure 8) including roads, waterways, and air transportation in freight shipping, all parties including logistics companies and the government had beneficial consequences in 2012 (Luis C. Blancas, 2014).

Figure 18. An illustration of the multimodal logistics system (Source:https://www.quincus.com/what-is-multimodal-logistics-how-do-you-optimize-it)

In Vietnam, it is unarguable that the managerial responsibilities of the Ministry of Transportation impact obviously on not only the implementation of multimodal logistics infrastructure development but the effective cooperation between provincial authorities and other government agencies as well (Luis C. Blancas, 2014). Vietnam Inland Waterways Administration, under the Ministry of Transportation, play an enormous role as the administrative agency of the inland waterways infrastructure and the logistics industry supervisor. As a result, they should stimulate a significant number of the cooperation projects in order that the private businesses are going to use the inland waterways logistics, and the managerial role of VIWA will be efficiently improved by reducing the fragmentation of their investment projects and optimizing the resources (Hoang Anh Dung, 2019).

4. Discussion of the feasible solutions and the further studies and for the development of inland waterways logistics systems

There is an overwhelming number of reasons which influence considerably the effectiveness of integrated inland waterways transportation system. Six determining factors that have a significant effect on the accomplishment of merchandise shipments through the inland waterways constitute the administration of the inland waterways ports, industrial areas and commercial businesses, cargo loading and unloading, asset inventory, in-stream overhauling, a convenient place for the periodic maintenance of inland waterways vessels, an ideal location for cargo fleets (Islam El-Nakib, 2006).

In addition, according to the report sponsored by the World Bank in 2014, the difficulties in inland waterways can be grouped by three complex divisions including the development of transportation infrastructure, the optimization of freight operations, and policy issued by many governmental agencies (Luis C. Blancas, 2014).

To optimise the annual budget for logistics and SCM through inland waterways, the first effective solution is that inland container depots should be located near inland water ports, and seaports, and as a result, there is the incentive for logistics enterprises to not only decrease their operational cost of using their truck fleets during rush hours but also encourage them to transport the commodity through inland waterways systems (Luis C. Blancas, 2014). Since 2014, there have been many inland container depots that were connected to inland waterways ports and seaports were invested and built in the focal economic zones in the South of Vietnam as figure 8.

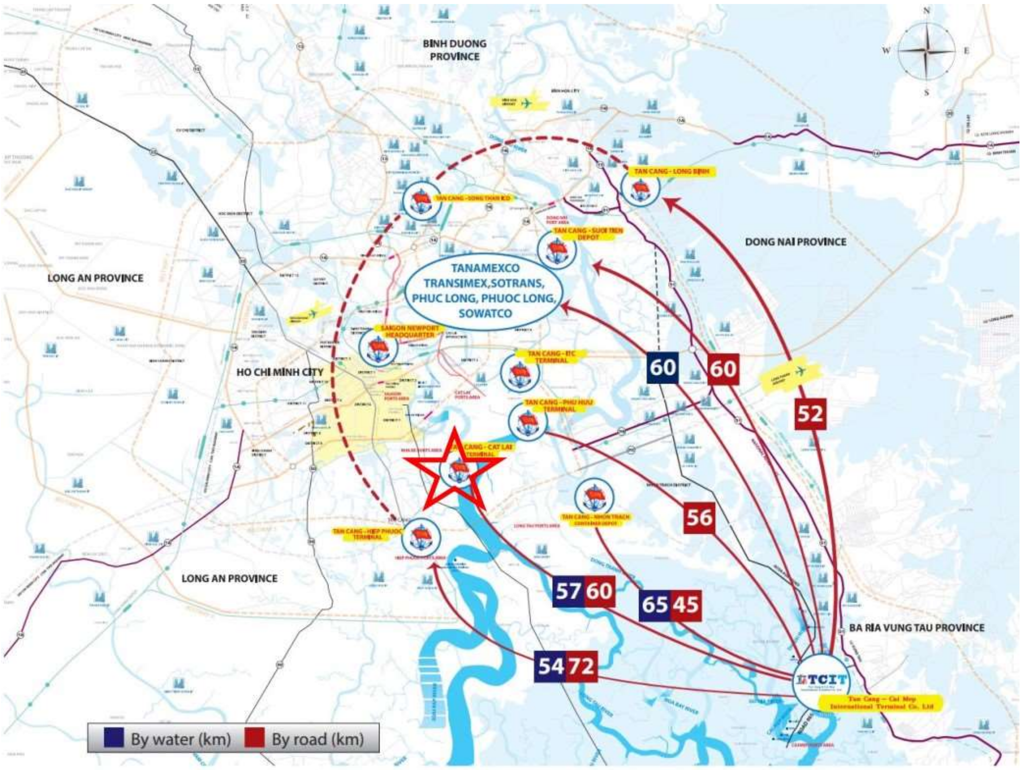


Figure 19. Tan Cang – Cai Mep international terminal and its logistics network
Source: Domestic Services - Tan Cang – Cai Mep International Terminal (tcit.com.vn)

The second workable solution is that the promotion of mutual cooperation and supports between inland clearance depots in Vietnam Logistics Business Association will not only increase the competitive strengths of the regional logistics system but also improve the freight operation services.

To foster the implementation of inland waterways shipping, the inland waterways sector development policies issued by relevant administrative agencies which are considered as a long-term solution should be based on market understanding and appraisal activities.

Regardless of the mentioned solutions, many aspects of integrated logistics and SCM should be considered and analysed in the ongoing researches, as follows:

- The productive balance of private and public businesses in logistics systems;
- Environmental issues in the operation of waterways traffic routes;
- The estimation of demand and supply based on the global, regional, local marketplaces;
- Well-qualified laborforce educated by academic institutes based on the practical demand;
- The effective coordination between three groups including administrative agencies, logistics enterprises, and universities in the promotion of integrated logistics solutions.

5. Conclusions

As can be analysed in the paper, it is essential to consider how to implement the advanced technologies in logistics and SCM in order that municipal governments can enhance their regional competitive strengths in the new circumstance caused by the Covid-19 outbreak. Since 2019, the detrimental impact of the covid-19 pandemic has affected not only regional economic and social development but also the process of logistics and SCM.

With regard to human resources in logistics and SMC, it is imperative that national, provincial, and local authorities should provide the adequate incentive to foster higher education institutes in opening the undergraduate and graduate majors in relation to logistics, SCM, and relevant occupations such as inland waterways captains, marine engineers, and customs officers, logistics professionals and as a result, they will satisfy the urgent demand of professional workforce for thousands of logistics enterprises in Vietnam.

In conclusion, the integrated logistics and SCM will be confronted with an overwhelming number of challenges in association with opportunities in the next decades, and hence Vietnam should tackle the ongoing problems and seize the reasonable chances to develop innovative methods and invest more inland ports (figure 10 and 11) for an effective integrated logistics systems associated with the comparative advantages of inland waterways transport.

Figure 20. SP-ITC International container terminal on Soai Rap river (Source: Screenshot from Google Maps)

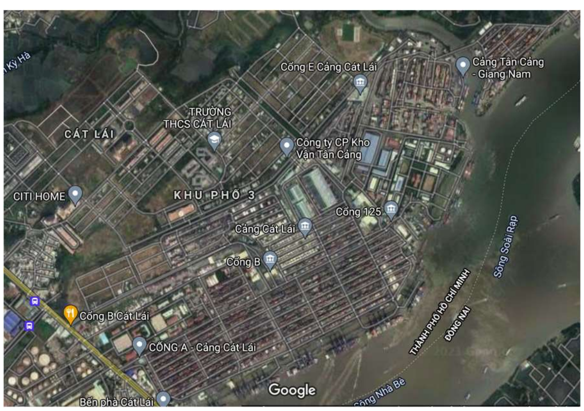


Figure 21. Cat Lai port on Nha Be river (Source: Screenshot from Google Maps)

References

Christopher, M. (2011). *Logistics and Supply Chain Management* (4th ed.). Great Britain: Pearson Education Limited. Retrieved 2021

Department of Information and Communication of Binh Duong. (2018, October 08). *Hiệp hội logistics tỉnh Bình Dương nhiệm kỳ 2018-2021*. Retrieved JULY 11, 2021, from http://stttt.binhduong.gov.vn/web/tin-tuc/1976?title=Hi%E1%BB%87p-h%E1%BB%99i-Logistics-t%E1%BB%89nh-B%C3%ACnh-D%C6%B0%C6%A1ng-nhi%E1%BB%87m-k%E1%BB%B3-2018-2021

Hoang Anh Dung, Y. Y. (2019). *Sustainable Development of Inland Waterways Transport in Vietnam: Strengthening the Regulatory, Institutional and Funding Frameworks.* Washington DC: The World Bank.

Islam El-Nakib, C. R. (2006). Challenges to inland waterways logistics development: the case of Egypt. *The 3rd International conference of Inland Water Transport (MELAHA 2006) Luxor*. Luxor, Egypt.

JSC, T. C.-L. (n.d.). Retrieved July 2021, from https://saigonnewport.com.vn/en/service/Pages/tan-cang-long-binh-icd-jsc.aspx?IDDichVu=4

Luis C. Blancas, J. I. (2014). *Efficient Logistics - A Key to Vietnam's Competitiveness.* Washington DC: International Bank for Reconstruction and Development / The World Bank.

Paul R. Murphy, J., & Knemeyer, A. M. (2018). *Contemporary Logistics* (12th ed.). Pearson Education Limited.

Stanley E. Fawcett, L. M. (2014). *Supply chain management from vision to implementation.* the United States of America: Pearson Education Limited.

Yin Lam, A. W. (2019, April 22). *Supporting Vietnam's economic success through greener, cheaper, and more efficient trucking*. Retrieved July 2021, from https://blogs.worldbank.org/transport/supporting-vietnam-s-economic-success-through-greener-cheaper-and-more-efficient-trucking

RESEARCH AND IMPROVEMENT ON WAREHOUSE OPERATION PROCESS BASED ON DISCRETE-EVENT SIMULATION MODEL

Nguyen Thi Phuong Linh[1] , Nguyen Huu Tho[2] and Do Ngoc Hien[1]

[1] *Department of Industrial Systems Engineering, Ho Chi Minh City University of Technology, VNU-HCM, 268 Ly Thuong Kiet, Ward 14, District 10, Ho Chi Minh City, Vietnam*

[2] *Department of Fundamentals of Mechanical Engineering, Faculty of Automotive, Mechanical, Electrical and Electronic Engineering, An Phu Dong Campus, Nguyen Tat Thanh University, District 12, Ho Chi Minh City, 729800, Vietnam*

Abstract. Logistics and supply chain management play an important role in today's economy. Logistics costs account for 20.8% of GDP. There are many waste problems in logistics activities. In which, materials handling costs do not create value but account for the majority of operating costs. Therefore, it is important for a business to solve these problems in order to achieve its goals of effective operation. This article will study this issue. From the beginning, modeling and simulation using FlexSim are realized. After that, the paper studies the outputs, finds bottlenecks and recognizes idle resources. Finally, find improvement scenarios to make the system operate more efficiently and give a reference for business.

Keywords: Warehouse; Simulation model; FlexSim; Supply chain.

1. Introduction

Nowadays, warehouses play a very important role in the production and business activities of enterprises, it even determines the success or failure of the firm. Warehouses play a critical intermediate role between supply chain members, affecting both supply chain costs and service (Faber, de Koster, and Smidts, 2013). In an effective supply chain management strategy, businesses have built efficient production management systems and integrated warehouse management systems in recent years. In the present context, supply chains are competing with each other mainly in terms of delivery lead time and overall product cost (Singh, Chaudhary, and Saxena, 2018).
In recent years, not many businesses can achieve the goal of effective warehouse management as required, especially small and medium enterprises (SMEs) and enterprises in developing countries- including Vietnam.

The need for a modern, efficient warehouse has never been more pressing. Labor costs, equipment costs, and customer demand continue to increase each year. Efficiency and optimization have become challenging tasks for enterprises, especially in the context of fierce competition and increasing integrated supply chains with reduced operating costs. Therefore, a question that businesses need to answer is how to effectively manage the warehouses in the current volatility production, management, planning, and supply chain. Usually all the required operations converting the raw material into product cannot be performed at one workstation; hence, materials handing is necessary. Although this activity generates place utility, it is not a direct operation; it adds cost but nit value. Hence reduced materials handling will increase productive efficiency. The amount of materials-handling activity in production processes is fairly large at present (Hitomi, 2017). The materials handling cost typically comprises 30-75% of the total cost. Effective materials handling can reduce this cost to 15-30% (Hitomi, 2017). The transfer and waiting times can be 95% of the time, only 5% is on the machine. Of that 5% of the total production time set up, inspecting, and idle times share 70%. Hence actual machining is only 1.5% of the total time (Hitomi, 2017).

Simulation is an effective technical mean for logistics system optimization. Simulation technique can show the status, identify the bottleneck of system and propose effective solutions to achieve the optimal resources. Flexsim is one of the most powerfully analytical tools for simulation. Flexsim is used for the simulation environment of discrete event system in manufacturing, material handling, logistics, transportation and management. Flexsim is a newly developed commercial discrete-event system simulation software by the United States Flexsim Software Production Company in recent years, it has a strong analytical capability and can be carried out according to different needs of simulation analysis provides a raw data input modelling, model simulation runs test, the results optimized to solve system integration of resources and optimization problems of main parameters. It is also a low-cost, fast and effective method (Peng, 2010; Zhang and Liu, 2006). By using simulation and modeling tool of FlexSim, this paper analyzes the warehouse operation process and builds a simulation model for the warehouse. Next, the performance of machines and workers in the current process will be analyzed. Finally find the bottlenecks of the system and free resources, then make improvements to improve the operational efficiency of the warehouse.

2. Modeling and simulation steps

1-Set targets: The first important thing to evaluate is to clearly determine the targets of the simulation study. Targets are achieved by evaluation criteria, by a system of questions that need to be answered.

2-Collect data: Collect basic data of the system. There are two aspects of data that the researchers should collect. First, data collected is based on the simulation goals aiming at keeping the system running normally. Second, the data includes starting-stopping conditions and variables of the system.

3-Conceptual model: Check the correctness of the model. The conceptual model must reflect the true nature of the system and the environment, but at the same time, it must be convenient and not cumbersome. If the conceptual model is not satisfactory, additional information and data must be collected to rebuild the model.

4-Build simulation model: Simulation models are programs that run on a computer. These programs are written in common or specialized programming languages

5-Running test: After installing the program, run a test to see if the model correctly reflects the system characteristics and correct programming errors.

6-Model validation: After the test run, verifying and evaluating whether the simulation model is satisfactory or not, if not, go back to step 2. Model validation and check conceptual model are two important procedures to confirm whether the model we build should work or not. Model validation checks whether the program can be run, whether the input and output data are convenient and correct. Conceptual model checking is an assessment of whether the model reflects the true nature of the real system, and whether the simulation results meet the requirements of the research.

7-Experiment Planning: At this step, define some simulation conditions. The first is to define the start condition, end condition, or run length. Next define the number of replication. Finally, determine the simulation time of each part or the whole model.

8-Experiment: Run the experiment according to the plan made in step 7. This is the step to perform the simulation. The output from this step is the output of the simulation.

9-Analyze the results: These results will be presented in an explicit form that is convenient to store and use. This step plays an important role in the simulation process.

10-Store the data: Use simulation results for the intended purpose and keep them as reusable documents.

Modeling simulation usually follows the basic steps as in Figure 1.

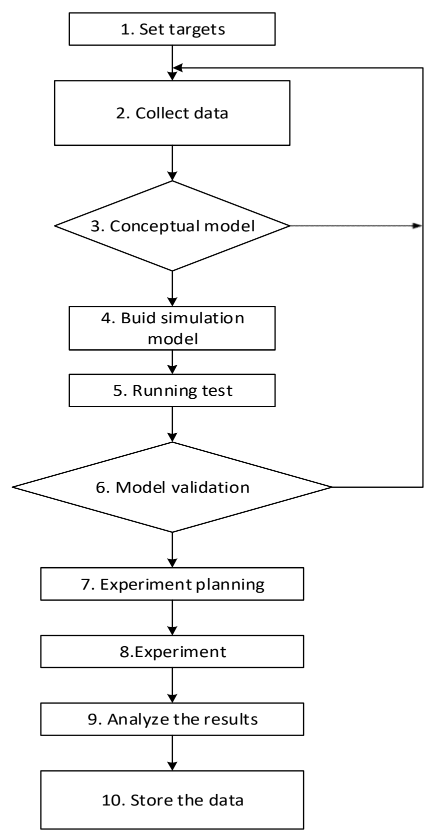


Figure 1. Procedure of simulation

3. Modeling for the Operation Process

3.1. The operation process and layout

The finished goods warehouse is where the shipping process is completed in a factory, not to mention the outbound operations of the shipment such as transporting the container to the port and loading the container onto the ship. So, basically, loading a truck or container is the last step in the shipping process at the factory. Furthermore, this paper only considers finished goods warehouses, so other warehouses are not considered in the paper. It is assumed in the model that everything before the production process has been done. That means the warehouse operation process mentioned in this paper starts from where the product is finished and approved by the quality control department. Next, the products are checked for classification and loaded onto conveyors for combining with a pallet. From this stage, the system will record the scanned products into the warehouse storage, and the inventory will increase the number of scans in the product. After the products undergo a complete scan, they are stacked to form a batch. Once the batching is done, the product is then moved to storage. The storage area is now separated into different areas to locate products based on the product type. After receiving the required order, the forklift will move the finished pallet to the unloading area. Scanners and sorters (separators in the model) will perform the process of separating pallets from packages for shipment.

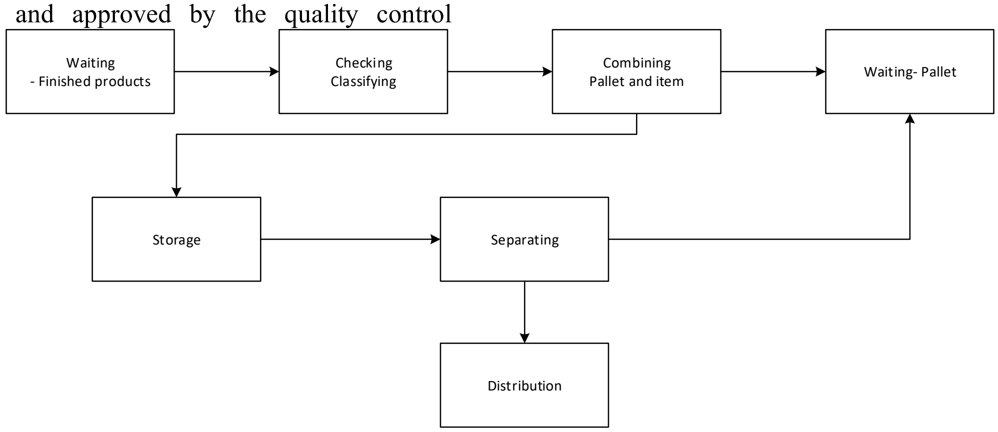


Figure 2. The operation process of warehouse

There are four main areas of the warehouse in this model including queuing area, sorting area, storage area, and loading area (see Figure 2). Each area consists of resources in which activities will occur that affect the status of products or their location. First, the queuing area is where the finished product appears and waits. In this area, workers will be responsible for loading and unloading goods to the inspection and classification area. Next, in the sorting area, there is a worker who checks the goods and sorts them to take to the storage area. The main part of the warehouse is the place to store the products, which is the storage place. In this area, products will be packed into batches and then stored on racks, and waiting for distribution. The final area is the distribution area, which uses equipment to load and unload goods to deliver to the user.

3.2. Simulation targets

The main goal of this paper is to reduce the average waiting time for products in a warehouse while making efficient use of resources. Other goals of this paper are to study the smart storage system, then define the bottleneck and establish an alternative scenario to improve the process. To evaluate the performance, several key performance indicators are given as follows.

- Total waiting time of a product, from completion of production (appearance) to completion of queuing (starting to be processed). After improvement, 30-40% of waiting time can be reduced
- Utilization of operational resources in the warehouse after improvement can be increased by at least 10%.

3.3. Data Collection

To achieve good results, we must collect system information and data. The warehouse in this model, as shown in Figure 3, operates 8 hours per day and operates for 5 days (144000 seconds). There are 3 types of incoming products and they are coded into 3 types of goods A, B, and C. Incoming products follow an exponential distribution (0.0, 11.26).

Collect data and use the Chi-squared test to check which distribution the station uptime follows, the results are given in the Table 1.

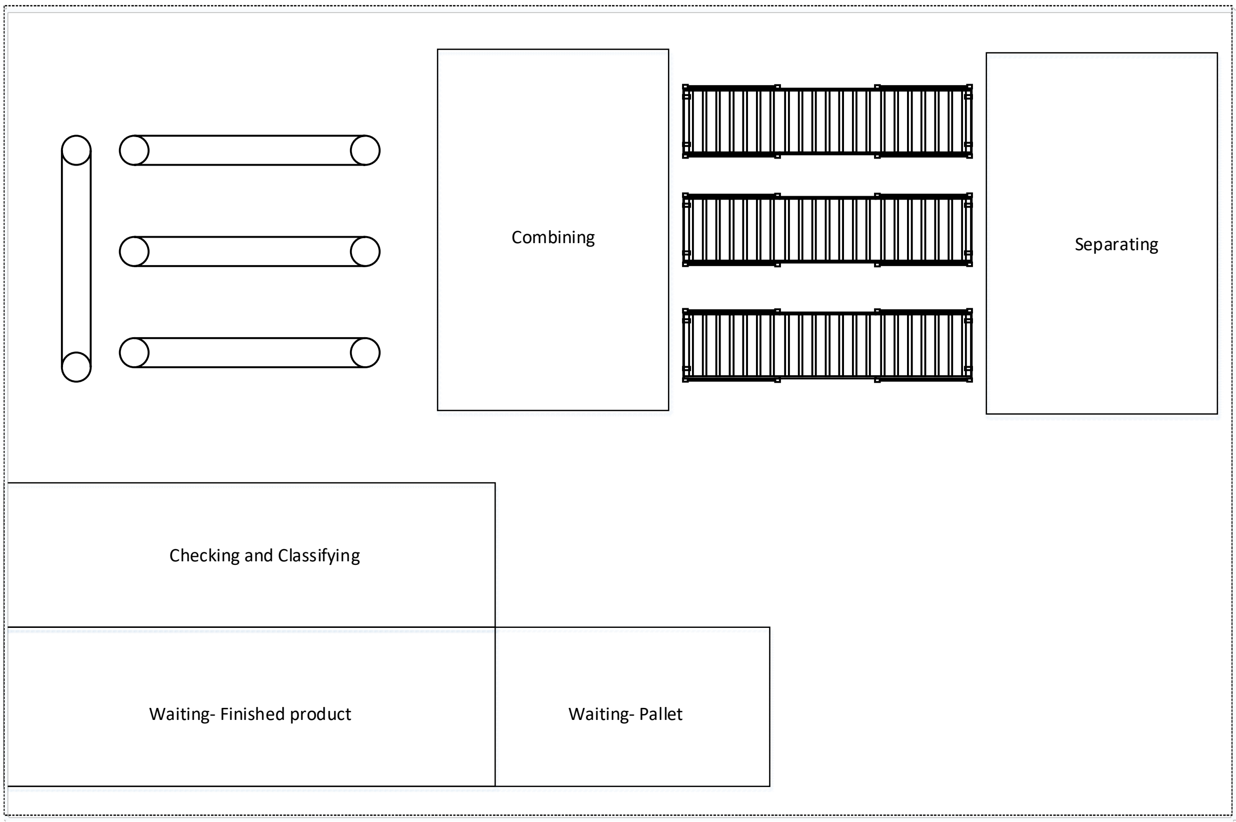


Figure 3: Warehouse layout

Table 1: Probability distribution of machines in the warehouse

Station	Machine	Distribution
Sorting	Processor	Uniform (15.9,20) (sec)
Combining	Combiner	Uniform (40.69, 59.86) (sec)

Table 2: Some parameters of machine, equipment and staff

No.	Object	Number	Speed (m/s)	Capacity
1	Conveyor	8	0.5	
2	ASRS	3	2	Max= 2 batches
3	Forklift	3	1	1 batch
4	Operators (not include forklift)	3	1.5	1 box , 1 pallet

3.4 Modeling for the Operation Process

Based on the analysis of the warehouse operation process in Section 3.1, this paper will build a simulation flowchart. This flowchart can be customized to suit the warehouse operation process based on the parameters in Table 2. The Figure 4 shows a flowchart that simulates the warehouse operation process.

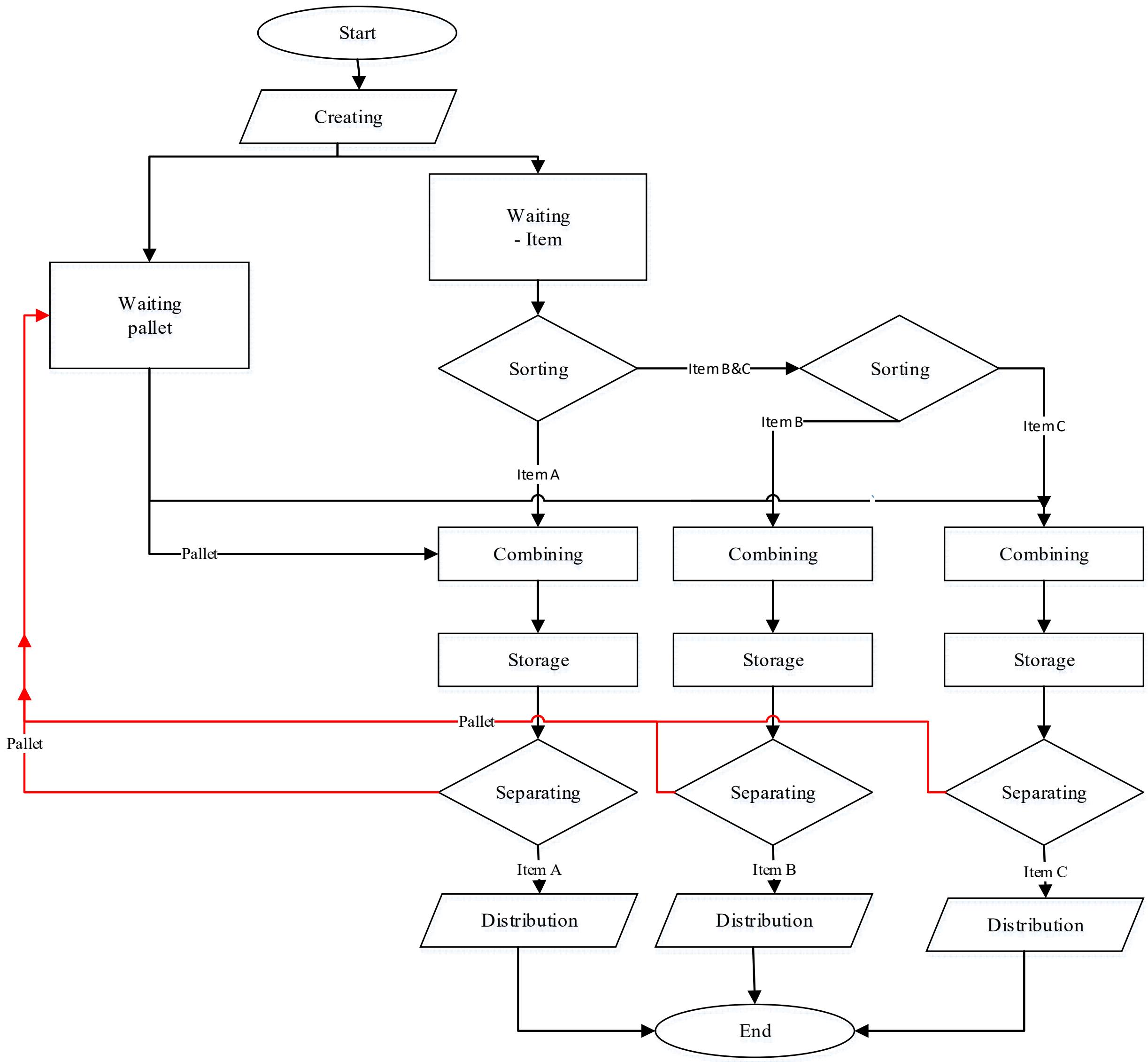


Figure 4. Simulation model for operation process

4. Simulation and Results analysis

4.1. Elements of Simulation and experiment

Table 3. The elements of simulation model

No.	Station	Object	Input	Remark
1	Queuing and sorting area	Source 1	Queue 1	Produce item A (red), B (green), and C (blue)
2		Source 2	Queue 5	Produce pallet
3		Dispatcher 1	Operator 1, 2	Coordinate operators to pick up items from queue 1 to processor 1
4		Queue 1	Processor 1	Wait
5		Processor 1	Conveyor	Check
6		Operator 1, 2	Queue 1 → Processor 1	Operators of picking from queue 1 to processor 1
7		Operator 5	Combiner 1,2,3	Operators of picking pallets
8		Conveyor	Combiner 1,2,3	Load items to combiner
9		Decision Points	Conveyor	In this model, these points are used to sort items
10		Combiner 1,2,3	Queue 6, 30, 29	Pack to batches
11	Storage area	Transporter 2	Queue 6,30,29	Pick up batches from combiner 1,2,3 to Queue 6,30,29
12		Rack 65,67,69	Separator 68,69,64	Store goods
13		ASRS 2,3,68	Rack 65,67,69	Pick up batches from Queue 6,30,20 to racks
14	Loading area	Separator 68,69,64	Conveyor	Separate items from batches
15		Sink 1,2,3	-	End- Place of Receipt

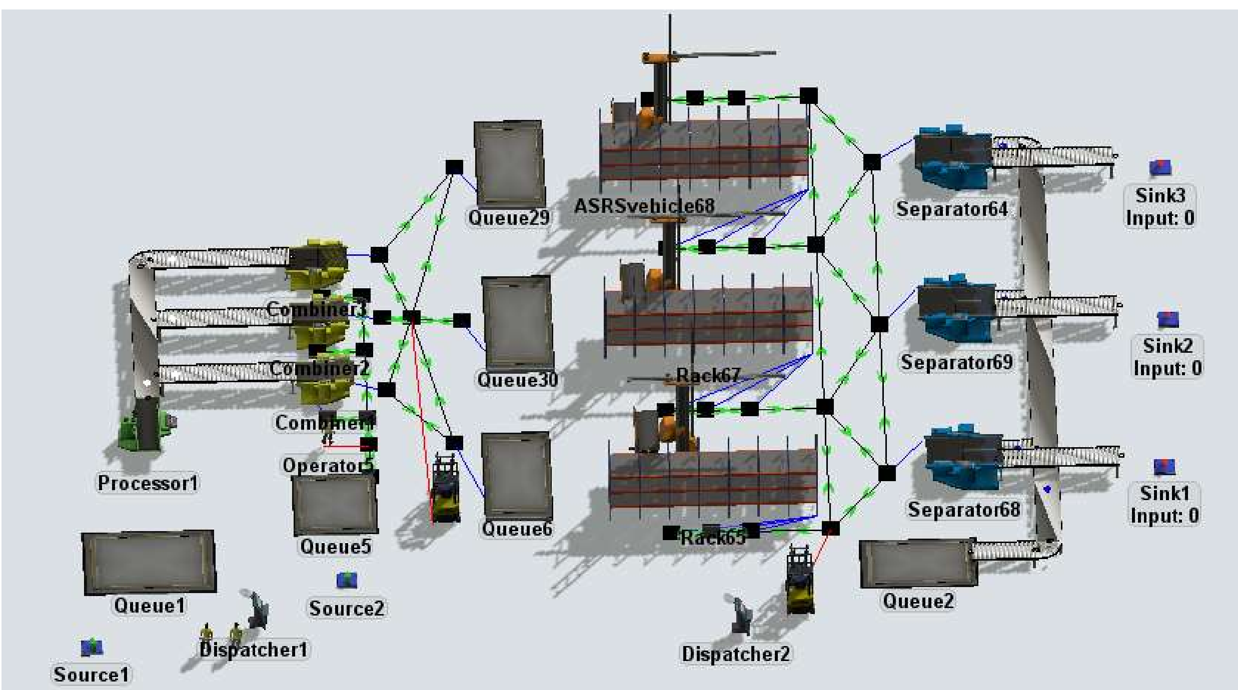


Figure 5: The simulation model

The elements of the discrete-event simulation model are presented in Table 3 and Figure 5. Experiment on simulation with the initial set number of replications is 10, it is determined that the time when the system starts to run stably is about 11000 seconds; therefore, the run length of the simulation will increase by 11000 seconds. From there, adjust the 5-day operating time of the simulation system to 155000 seconds.

To increase the reliability of the simulation, the calculation of the number of simulation replications is performed based on the simulation results with the initial number of iterations of 10.

$$n \cong n_o . \frac{h_o^2}{h^2} \quad (1)$$

where, h_o is one-side width of initial confidence interval and h is one-side width of desired confidence interval.
Because the result n = 0.0192 < 10 can be calculated, the number of replications is small. Therefore, the number of hypothesis replications n_o = 10 should be chosen as n = 10.

4.2 Output and Analysis of simulation results

After running the simulation with 10 replications, the results for average queue time is shown as Figure 6. The waiting time is 26486 seconds (more than 7 hours).

Staytime - Average Average

	Mean (95% Confidence)					Sample Std Dev	Min	Max
Scenario 1	26445	<	26486	<	26528	58	26377	26559

Figure 6. Output result of waiting time

In 5 days, the source generated only 42.59% of the goods that appeared in the queuing area (see Figure 7). That means that the receiving capacity in this area is overloaded and 57.41% of the goods are locked from the production area and have not been put into the warehouse.

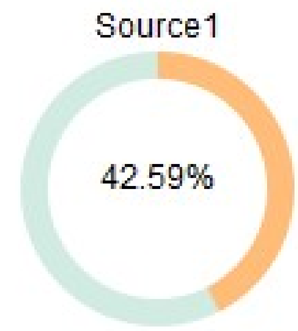


Figure 7. The generating percentage from source

Look at the performance of operators, we see that operators 1 and 2 are only efficient 21.5% of their working time. Operator 5 is only 1.89% efficient (see Table 4). Thus, the operators were really inefficient, with too much idle time, while the number of goods in queues was still overloaded.

Table 4: The utilization of operators

Object	Travel loaded %	Idle %	Travel Empty %
Operator 1	9.86	78.50	7.57
Operator 2	9.86	78.50	7.57
Operator 5	0.81	98.11	0.77

Conveyors are operating at less than 20% capacity and this level of utilization is inefficient. The processor works 56.97% as in Figure 8 and Table 5, this efficiency is relative, not high; however, using only 1 machine makes the goods stagnate in the queuing area quite a lot.

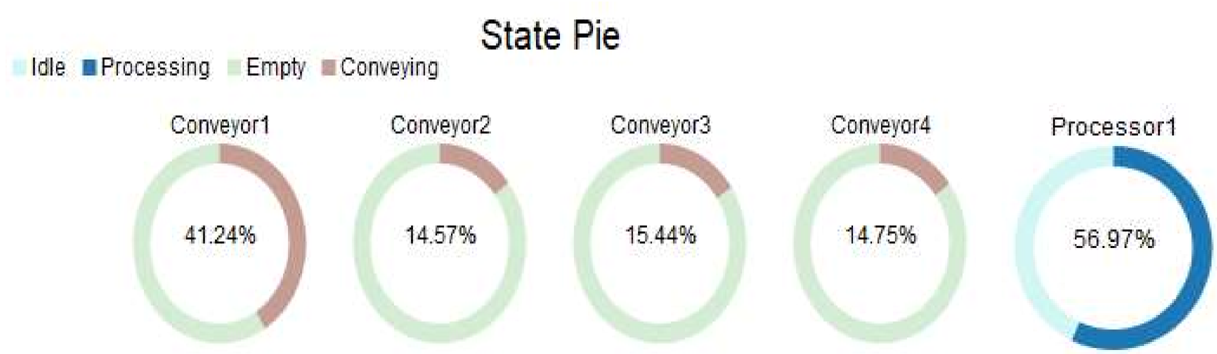


Figure 8: The utilization of conveyors and processor

Table 5. The utilization of conveyors and processor

Object	Conveying %	Processing %	Idle %
Conveyor 1	41.35	-	58.65
Conveyor 2	14.57	-	85.52
Conveyor 3	15.44	-	84.55
Conveyor 4	14.75	-	85.19
Processor 1	-	56.97	43.02

Looking at the above results, we come to the conclusion that the queuing time is too long (7 hours) while the resources are not used efficiently. It is necessary to find ways to improve to reduce waiting times and increase the efficiency of the use of resources, thereby improving the performance of the warehouse.

4.3 Optimization of Simulation model

Overloaded items in the queue. The bottleneck can be solved by increasing the number of vehicles. The utilization of resources is low.

Variables that affect the operational efficiency of the warehouse include the number of operators, machines, racks, and forklifts. The limit the number of these resources based on the lowest and highest costs of the business as well as the design space of the warehouse. They are shown in Figure 9. The main objectives are to maximize the utilization, minimize the queuing time, and minimize operators.

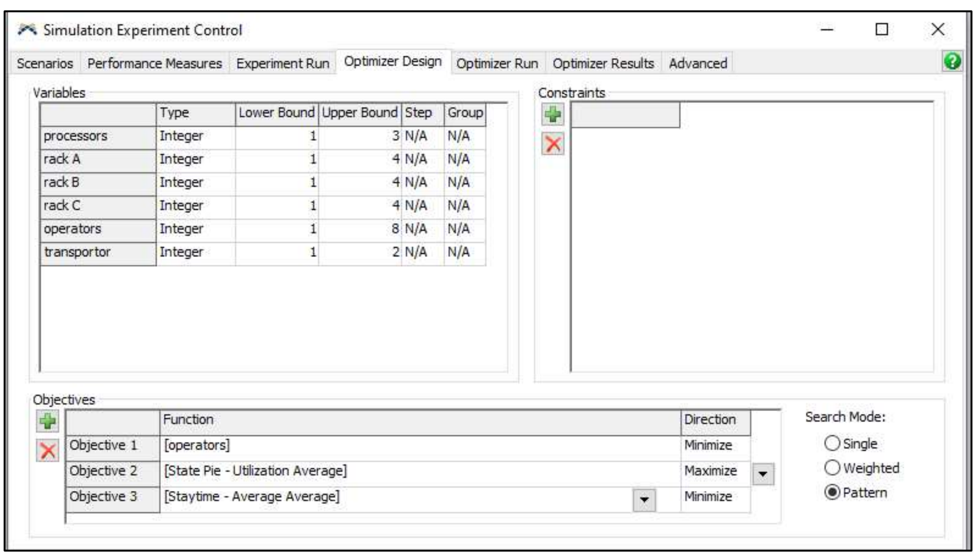
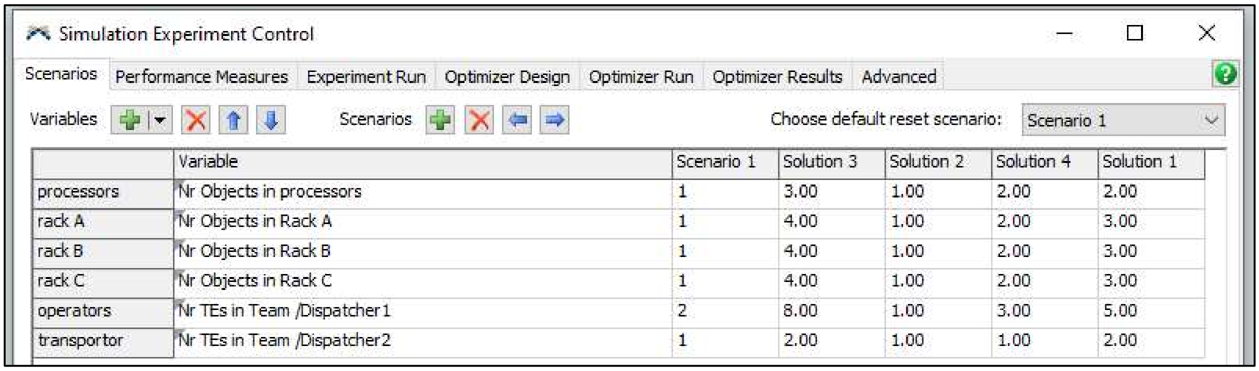


Figure 9. Optimizer design

The results after optimizing are shown in Figure 10.



Simulation Experiment Control

Scenarios | Performance Measures | Experiment Run | Optimizer Design | Optimizer Run | Optimizer Results | Advanced

Variables | Scenarios | Choose default reset scenario: Scenario 1

	Variable	Scenario 1	Solution 3	Solution 2	Solution 4	Solution 1
processors	Nr Objects in processors	1	3.00	1.00	2.00	2.00
rack A	Nr Objects in Rack A	1	4.00	1.00	2.00	3.00
rack B	Nr Objects in Rack B	1	4.00	1.00	2.00	3.00
rack C	Nr Objects in Rack C	1	4.00	1.00	2.00	3.00
operators	Nr TEs in Team /Dispatcher1	2	8.00	1.00	3.00	5.00
transportor	Nr TEs in Team /Dispatcher2	1	2.00	1.00	1.00	2.00

Figure 10. Results of optimization

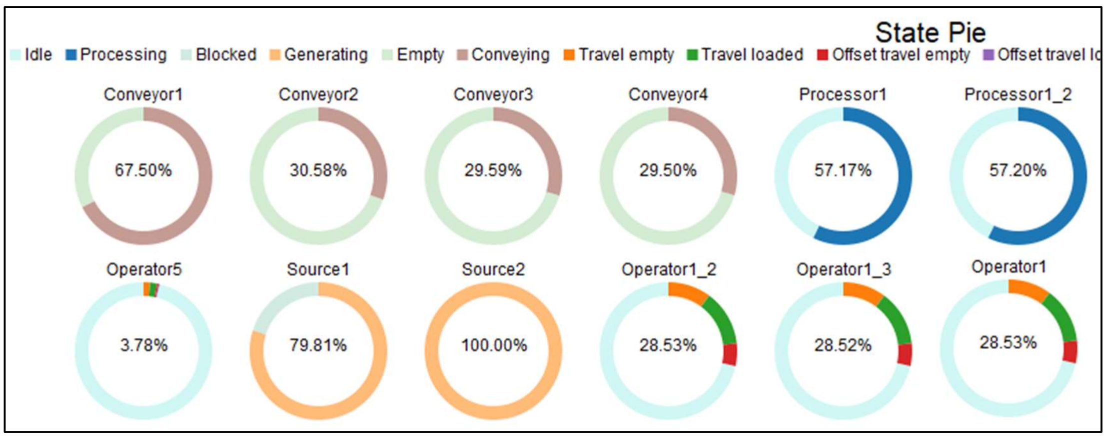


Figure 11. The utilization of resources

Run simulation with 10 replications and then shows the results along many solutions. The solution 4 gives the results agreeably. Solution 4 as in Figure 10, the number of resources increased to 2 processors, 6 racks,

3 operators; therefore it costs less than option 3, cuts 3.5 hours (achieve target) and usefulness increases by 11% (see Figure 11 and Table 6).

Table 6. Contrast of two simulation results

Evaluation Index		Before Optimization	After Optimization	Contrast
Waiting time – (Queuing time) (s)		26486	13118	−13368
Utilization (%)		25.675	36.183	10.508
Processors (%)	1	56.97	57.17	0.2
	1.2	-	57.20	-
Operators (%)	1	21.5	28.53	7.03
	2	21.5	28.52	7.02
	1.2	-	28.53	-
	5	1.89	3.78	1.89
Conveyors (%)	1	41.35	67.5	26.15
	2	14.57	30.58	16.01
	3	15.44	29.59	14.15
	4	14.75	29.5	14.75
Rack		3	6	3
Source (%)		42.59	79.81	37.22

5. Conclusions

In the current context, not all businesses can achieve the goal of effective warehouse management as desired, especially small and medium enterprises (SMEs). The labor costs, equipment costs, and customer demand continue to increase each year. There are many activities inside the process, which do not directly create the product but cost and create no added value, such as materials handling costs. Therefore, cutting materials handling costs will increase operational efficiency in the warehouse

Simulation modeling is an effective tool to evaluate the overall system over a long period of time, giving researchers insight into the operation of a system, which can change operating principles, resources to improve system performance, test new systems before commissioning or make improvements, and obtain information without affecting the actual system. The Flexsim simulation tool helps researchers have a more intuitive and dynamic view of the system's state and easily identify the system's bottlenecks.

This presented the contributions as follows: Survey and identify bottlenecks and resource usage in the current warehouse; Modeling simulates operating status for 5 days with an average waiting time of up to 7 hours with only 2 operators and overall resource efficiency of 25.675%; Then propose an improvement plan to reduce the average waiting time to 3.5 hours and increase the resource utilization to 36.183% through the Flexsim simulation tool.

Moreover, this paper shows how process simulation implementation, especially FlexSim, can be very useful for small and medium industries. Scientifically, this study sheds light on the reality of warehouse operations, waiting time too long and not using resources will increase costs for businesses. In practical terms, the research results will contribute to helping businesses make decisions to improve existing processes, helping logistics and supply chain management professionals focus on cutting costs of material handling activities create no value and choose a strategy that works.

Acknowledgements

We would like to thank the FlexSim company for providing the academic license to complete our simulation model.

References

Faber, N., de Koster, M. B. M., & Smidts, A. (2013). Organizing warehouse management. *International Journal of Operations & Production Management, 33*(9), 1230-1256. doi:10.1108/IJOPM-12-2011-0471

Hitomi, K. (2017). *Manufacturing Systems Engineering: A unified approach to manufacturing technology, productionmanagement, and industrial economics*: Routledge.

Peng, J. (2010). *Optimal Allocation in Automatic Storage & Retrieval System Based on Flexsim Simulation.* Paper presented at the 2010 International Conference on Internet Technology and Applications.

Singh, R. K., Chaudhary, N., & Saxena, N. (2018). Selection of warehouse location for a global supply chain: A case study. *IIMB Management Review, 30*(4), 343-356. doi:https://doi.org/10.1016/j.iimb.2018.08.009

Zhang, X., & Liu, Y. J. B. T. U. P. (2006). Practical tutorial of system simulation software about Flexsim 3.0.

WORK DESIGN AT STEEL MANUFACTURING TO IMPROVE SAFETY OF PREVENTIVE MAINTENANCE

Luong Thanh Tri[1], Nguyen Tan Thanh[2], Nguyen Anh Kiet[3], Nguyen Phuong Nguyen[4], Do Ngoc Hien[5]

Industrial Engineering Major, Ho Chi Minh City University of Technology – VNU

Email: [1]lttri.sdh20@hcmut.edu.vn, [2]ntthanh.sdh20@hcmut.edu.vn, [3]nakiet.sdh20@hcmut.edu.vn, [4]nguyen.nguyen1999@hcmut.edu.vn, [5]hienise97@hcmut.edu.vnc

Abstract. Nowadays, safety first especially in heavy industries such as steel making production is paid much attention. The preventive maintenance of equipment and machinery in steel companies has faced with high risks and taken a lot of working time. The paper discusses the safety concerns on the preventive maintenance for roller motor at steel rolling station. Safety problems causes accidents such as hand injury, falling motor in legs, back pain and so on by the rudimentary supporting tools and confined space. As a result, the risk of co-worker accidents would increase if carelessness occurred. The DMAIC (Define – Measure – Analyze – Improve – Control) process would be applied. Work design and time study methods would improve the safety and efficiency of the preventive maintenance. The research brings benefits not only on minimizing the risk of working processes but also reducing unnecessary resources.

Keywords: Work design; DMAIC method; Safety; Lean; Ergonomic.

1. Introduction

Worker who regularly work in factory, affected by potential hazards about safety and labor health. Impacts on worker include fatigue, pain, illness, low morale, low production, frustration. Therefore, workplace ergonomics is becoming more important for manufacturing environments. Effective ergonomic strategies can increase productivity, reduce work injuries, and improve workstation design and layout. Ergonomics can support lean transformation and lean transformation can lead to ergonomic risk reduction (Testani, 2013). This study, conducted at a steel manufacturing company, incorporates a lean - ergonomic approach to the steel rolling station for motor roller VR43 maintenance.

Lean Manufacturing (LM) is a production method that was first introduced by the Japanese company Toyota (Khani, 2018). LM has been studied and applied in a wide variety of companies with positive results in production and administrative processes, but much more research remains to be done on the application of this philosophy to Occupational Safety and Health (Madariaga, 2013), (Carreras, 2010). According to different occupational health studies, muscular-skeletal diseases are the main cause for morbidity among the working population, producing lost days of work and decreasing business productivity (Nieuwenhuijsen, 2010). In addition to other risk factors such as over-exertion, repetitive motions, and inadequate postures, which are often the result of poorly designed and inefficient work areas and bad practices, and which may be improved by means of LM (Testani, 2013).

Assessment of ergonomic risk factors in the workplace should be carried out continuously. Ergonomic risks exist in any industries, and hence companies should be prepared to prevent, reduce the risks and their impact. Workplace safety and lean are highly interrelated (Saurin, 2006). Workplace ergonomics and lean manufacturing are highly inter-related. Lean focuses on eliminating waste in the process, while ergonomic safety focuses on eliminating hazards of injury to workers. However, safety risks (injuries and illnesses) can lead to lean wastes and vice versa. Improving a safe workplace can reduce operator stress and fatigue, while improving their productivity and reducing injuries.

2. Literature review

A lot of research has focused and implemented tools and methodologies to develop strategies to reduce risk factors for occupational accidents. Worldwide, 270 million work accidents (WA) and 160 million occupational diseases (OD) are reported each year (Leão, 2019). In Colombia, according to the Colombian Federation of Insurers (Fasecolda), there were 723,836 work accidents and 9,583 occupational diseases in 2005 (Federación de Aseguradoras

Colombianas Fasecolda). Therefore, companies need to implement programs to prevent risks to occupational safety. Properly implemented risk management and timely prevention contribute to reducing accidents and work-related illnesses. Besides, companies must work hand in hand with insurers and employees, involving them to participate in training to further reduce work-related events (Briceño, 2003).

Below we prodvice a literature review on studies that employ lean methods to reduce accidents and occupational diseases, these methodologies help to eliminate waste and make process leaner.

Hazard Identification and Risk Assessment (HIRAC) was done on paper pleating production line that assembled air filter product for automotive industry. Based on the assessment, the cutting process had the highest risk score value that triggerred for improvement as to reduce the risk level (Ahmad, 2017). Poka-yoke concept, which was considered as one of the engineering control methods, was used to identify the root cause of the problem. The solution of implementing a new cutting tool design significantly reduced the risk by more than 55%. The new process also made the pleating process became easier to handle; hence, it increased the production linearity and improved the product quality. In addition, (Geraldo dos Santos, 2015) automotive company applied some methods of analysis process by the correlation with implementation of LM and working conditions ergonomics. In which refers to the results (Quantitative - same volume of products with the lowest number of employees) and the (Qualitative - faster, accuracy, timeliness, and improved quality of the product). This study demonstrates a result of analysis of objectives achieved by continuous improvement applied between the elimination of waste and increased productivity and the impacts, the ergonomic conditions to carry out operations and consequently the results achieved in improving the productivity and well-being of its employees.

LM tools were orginally developed to improve manufacturing production systems. For this reason, they are used more frequently in manufacturing. But LM also can may apply for orther industry. As (Tortorella, 2020) The Ergo - Lean approach applies Lean Manufacturing to the field of Occupational Safety and Health, focusing on reducing activities that do not add value and on reducing the associated muscular-skeletal disorders. The purpose of this study is to integrate Lean Manufacturing tools in risk management to help control work-related accidents. Based on work accident data at a service company, the causes of the accidents were analyzed and the main Lean Manufacturing tools were reviewed and integrated, to develop a new methodology for the implementation of these tools to help prevent work - related accidents and diseases and to develop a culture of self-care.

3. Methodology

The DMAIC (Define-Measure-Analyze-Improve-Control) method in Six Sigma is often described as an approach for problem solving (Mast, 2012). But this study follows the DMAIC process using the Lean – Ergonomic mehods, as an analysis was performed of find the root cause that affects safety at the steel rolling station by using tool such as FMEA method (Failure Mode and Effects Analysis) and Pareto chart. Conduct time measurement for roller motor VR43 maintenance by timing each manipulation, describe process again by process chart. Subsequently, research proposes improvement for concern safety. Lastly, improvement control will be exercised by design work and make process leaner.

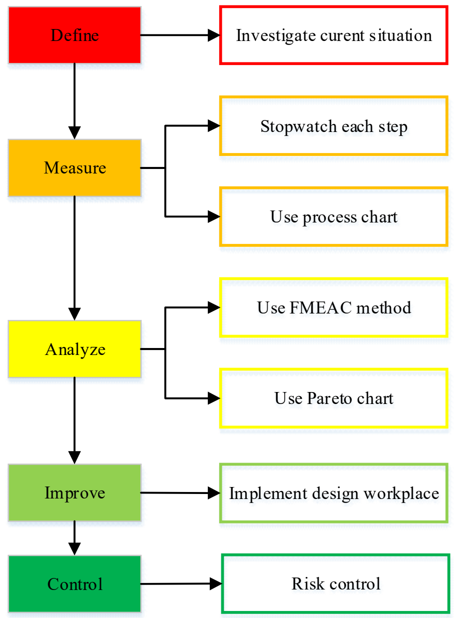


Figure 1. Implementation process

- Defining the goal and its requirements:

In this first step, it is necessary to define what the problem is. It is important to recognize and identify the following factors:

✓ Who is the subject?
✓ What are the key stages in this process?
✓ Functions of the process;
✓ The actual state of the matter

- Measurement the current process:

The purpose of this step is to establish the most important aspects of the current process and to collect relevant data. The important aspects in this step are:

✓ Identification of valid and reliable metrics,
✓ Checking if there is enough data to measure,
✓ Documentation of current performance and effectiveness,
✓ Performing comparative tests.

- Analysis

Then, the data collected in the previous stages will be analyzed. The purpose of this step is to determine the root cause relationship and find the reason for the defects and errors. The basic tools are used to:

✓ Determine the gap between current performance and desired performance
✓ Define input and output
✓ List and prioritize potential opportunities for process improvement
✓ Identify errors and causes using Ishikawa diagrams.

- Improvement

Existing processes will be improved using innovative techniques and solutions. Brainstorming sessions can be a useful tool. In addition, businesses can use other solutions such as:

✓ Creative ideas;
✓ Focus on the easiest and simplest solutions;
✓ Create a detailed execution plan;
✓ Implement improvement measures.

- Control

This step focuses not only on control but also on monitoring. This stage ensures that any deviations can be corrected in the future. Monitoring leads to sustainable improvements and long-term success. Therefore, regular monitoring needs to be taken seriously.

4. Case study

4.1 Defining the problem

At the steel rolling station, the task of roller motor VR43 is to transport steel from the area behind the C41 cutter to the cold floor. The frequency of replacing and removing the roller motor at this station is of a regular and periodic nature with the causes of replacement including 3 main points:

- Periodic maintenance
- The roller is damaged, worn
- Broken during operation, the motor coil was burned, the bearing was broken, the shaft was broken during operation

The investigation of current situation shows that the first is an area with many different terrains, the second is a crowded operation area, the third is dismantling and assembly process is performed by manual in the current, the fourth is easy to happen. Loss of occupational safety and health such as finger jams, head touch, tripping, motor slippage, back pain, heat, and noise. Such current situation adversely affects the safety, progress of the work as well as the mental state of the operator.

Figure 2. Steel rolling station

Figure 3. Roller motor VR43

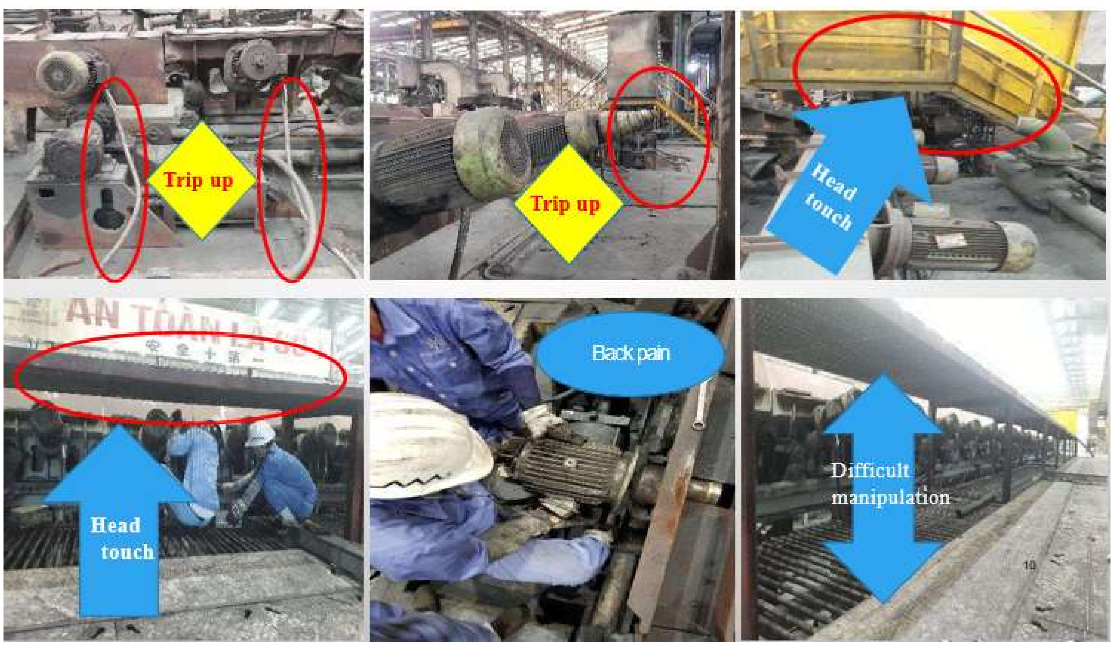


Figure 4. Some unsafe images at steel rolling

4.2 Measuring data

This research performs 10-step time measurement, measuring 30 times in each step and averaging. The stopwatch results are described through the process chart below in Tab. 1:

4.3 Analyzing the main issue

As a proactive risk analysis technique, the failure mode and effect analysis (FMEA) has been widely utilized to assure the reliability and safety of various fields. However, the conventional FMEA model possess several drawbacks such as overlooking FMEA team member's psychological behavior, uncertainty, and interaction relationships among risk indicators (Rao, 2020). In this paper, we link to FMEA (in Tab. 2) analysis and Pareto chart to show which manipulation are not safe for the operator. This study reviews the current operation process and analyzes the risks of each action through the evaluation of the indicators of the FMEA method. The RPN can be easily calculated by multiplying three numbers. All 3 indicators are rated on a scale of 10.

- *Severity*: the higher the score, the more severe it is, and vice versa
- *Occurrence*: The higher the score, the more it appears, and vice versa
- *Detection*: the higher the score, the harder it is to detect, and vice versa

Table 1. Process chart

Step	Process	Operation	Movement	Inspection	Delay	Storage	Thời gian (s)
1	Remove the power source from the roller motor	○	➡	◗ (open)	□	▽	114.2
2	One person removes four screws to fix the roller motor while two people use the lever to support the motor to avoid falling on the leg	●	⇨	◗ (open)	□	▽	127
3	Use a hammer and split knife to gently knock the motor apart from the fixed chute	●	⇨	◗ (open)	□	▽	39.2
4	Outward the roller motor with the lever	○	➡	◗ (open)	□	▽	73.9
5	Two people lift the engine and put the engine in the third position holding the m12 x 50 screw to prepare it to be mounted on the face	○	➡	◗ (open)	□	▽	65
6	Two people use levers to lift up the position to be installed	●	⇨	◗ (open)	□	▽	44.4
7	Two people hold the two sides of the swingarm, the third person aligns the motor flange to the fixed flange adopts 1 round steel rod guide pin	●	⇨	◗ (open)	□	▽	612.7
8	Use a spanner to tighten the four screws fixing the flange	●	⇨	◗ (open)	□	▽	174.9
9	Use your hand to gently crank the roller to see if it will be rubbed	○	⇨	◗ (filled)	□	▽	41.3
10	Check direction and inspection before power source reconnection	○	⇨	◗ (filled)	□	▽	317.5
						Total Time	**1610.1**

Table 2. FMEA analysis

Process	Step	Manipulation	Risk	Labor (people)	Serverity	Occurrence	Detection	RPN
Dismantling process	1	Remove the power source from the roller motor	Hand injury	3	3	1	3	9
	2	**One person removes four screws to fix the roller motor while two people use the lever to support the motor to avoid falling on the leg**	**Hand injury, motor falls to foot, back pain**		**8**	**6**	**6**	**288**
	3	Use a hammer and split knife to gently knock the motor apart from the fixed chute	Hand injury		3	2	5	30
	4	**Outward the roller motor with the lever**	**Hand injury, motor falls to foot, back pain**		**9**	**8**	**7**	**504**
Assembly process	**5**	**Two people lift the engine and put the engine in the third position holding the m12 x 50 screw to prepare it to be mounted on the face**	**Hand injury, motor falls to foot, back pain**		**8**	**7**	**8**	**448**
	6	Two people use levers to lift up the position to be installed	Hand injury, body pain		4	2	3	24
	7	**Two people hold the two sides of the swingarm, the third person aligns the motor flange to the fixed flange adopts 1 round steel rod guide pin**	**Hand injury, motor falls to foot, back pain**		**7**	**5**	**8**	**280**
	8	Use a spanner to tighten the four screws fixing the flange	Hand injury		2	3	2	12
	9	Use your hand to gently crank the roller to see if it will be rubbed	Hand injury		1	5	2	10
	10	Check direction and inspection before power source reconnection	Hand injury		1	6	2	12

Current process review of the VR43 roller motor dismantiling and assembly process consists of 10 steps, step 1 to step 4 is the dismantling process, step 5 to step 10 is the assembly process. Applying the FMEA and Pareto method in the figure below, we find that the unsafe two-step dismantling is in step 2 and 4, the assembly process also has two unsafe steps located in step 5 and 7.

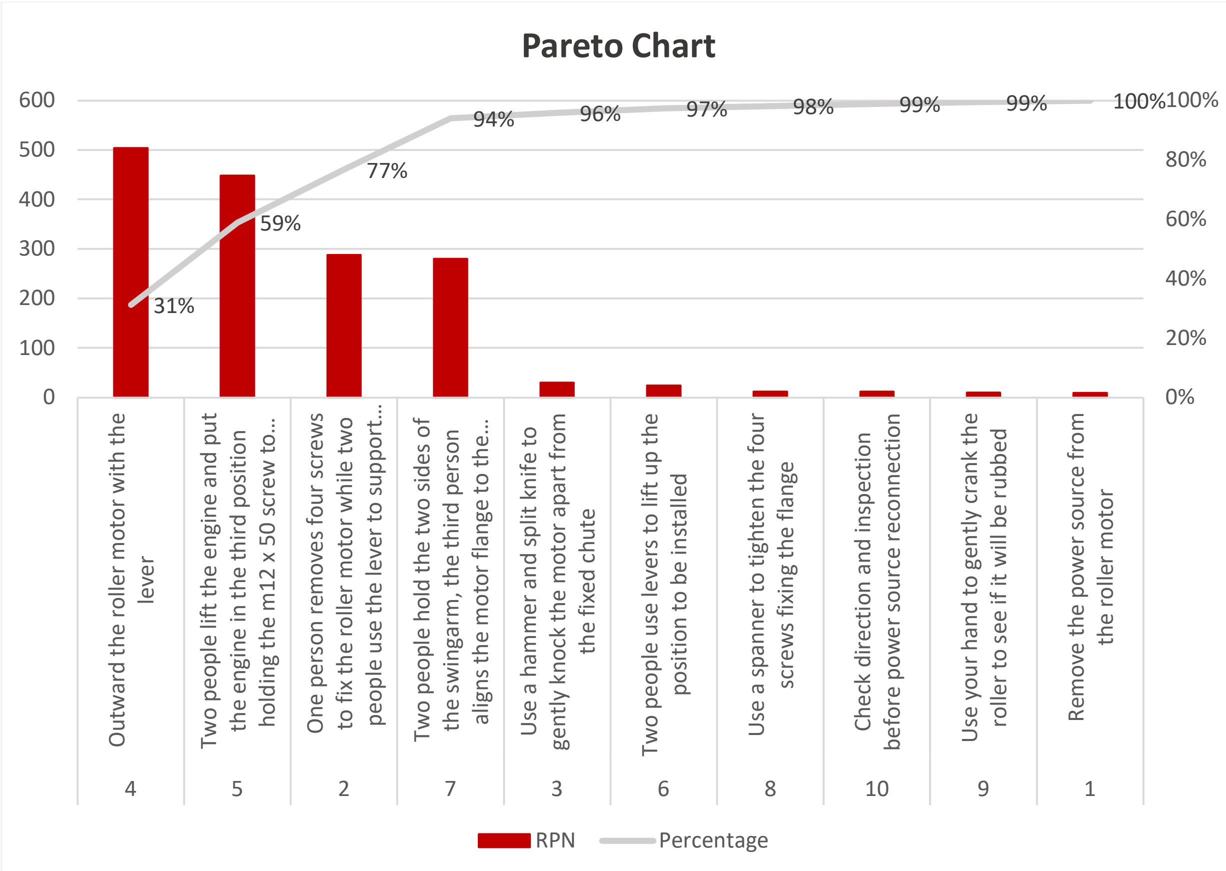


Figure 5. Pareto chart

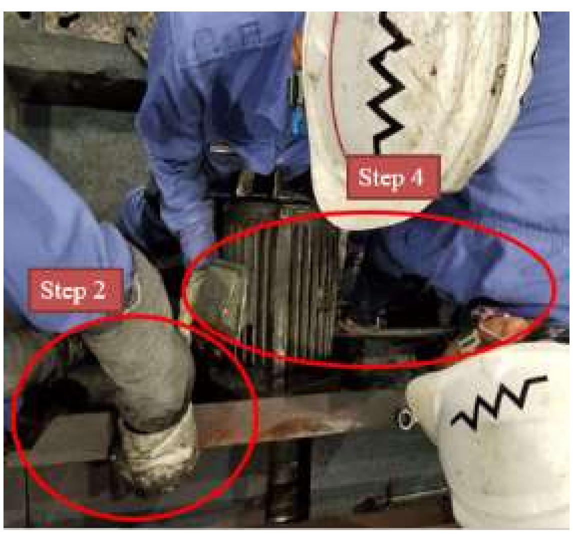


Figure 6. High risk in dismantling process

Figure 7. High risk in assembly process

4.4 Improvement proposition

From the knowledge and information from previous phases, the best possible solutions are identified, actions implemented and tested for validation (Rao, 2020). Four steps in the dismantling and assembly process are identified as occupational safety hazards leading to hand injurin, back pain, motor falling on the leg... To improve that problem, replace the bolt normally (C45) hexagonal head with studs (stainless steel 304) at the top 2 positions and the implant bolts will be firmly fixed to the VR43 holder to avoid stopping the guide screwdriver causing danger that can drop the motor and reduce get resources. Using implanted bolts makes it easier to put the motor in place than before because 2 implanted bolts are also responsible for guiding pins. Using 304 stainless steel material for implant bolts increases the resistance to high temperatures in the area, as well as the rigidity to ensure long-term durability when the manipulation work is periodic and frequent, the design must also ensure guarantee in the process loosen the nut, the implant bolt must also stay in place.

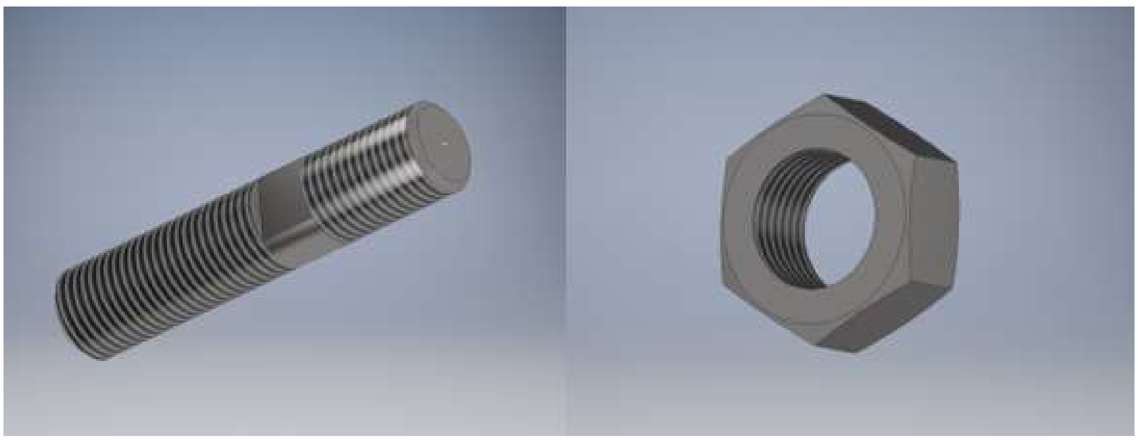


Figure 8. 3D drawing of implant bolts

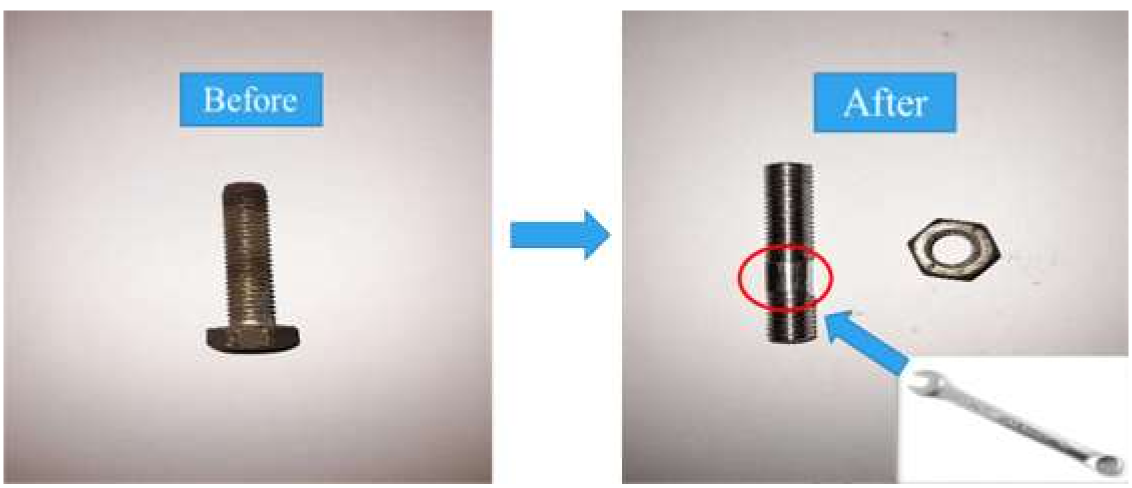


Figure 9. Pictures of normal bolts (left) and implant bolts (right)

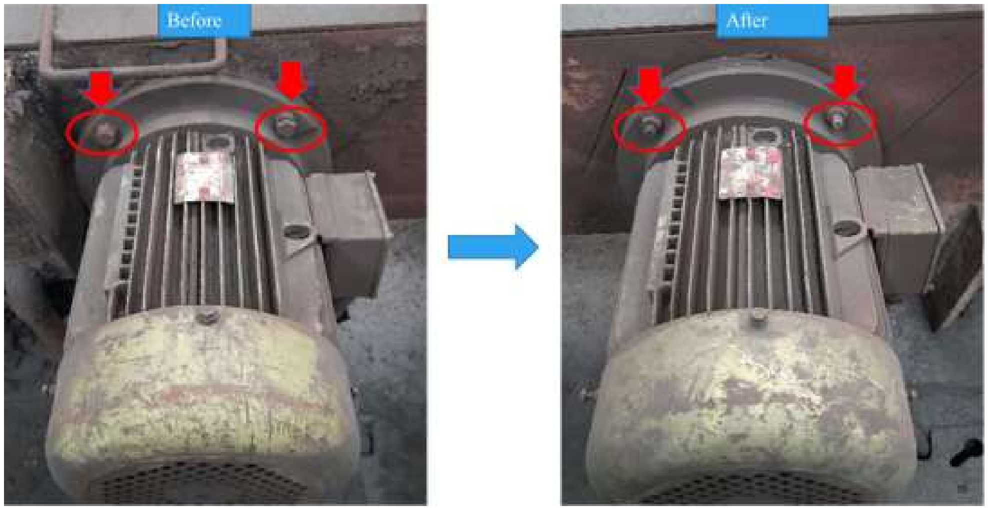


Figure 10. Photos before and after replacing implant bolts

Figure 11. The picture shows that there is no longer a hazard when replacing implant bolts

Figure 12. Replacing implant bolts makes it easier for the operator

Just replacing the implant bolts reduces 1 resource, but the other 2 operators still must lift and hold the motor, which can cause hand pain and motor fall to the feet. Therefore, the research continues to design the trolley to fix this motor to avoid the 2 operators lifting and holding the motor for too long.

Figure 13. 3D drawing of trolley

Figure 14. Experimental images

4.5 Contiuous control

After implementing the changes, they should be controlled to check if they influence the production process positively and bring any profits to the company (Smętkowska, 2018). For this study, the test results were measured through the process chart, showing that the results were overcoming occupational safety hazards, helping the operator to work comfortably and feel safe. more complete. In addition to overcoming safety risks and bringing ergonomics to operators, Lean has helped reduce operating time 18% compared with originall, reduce resources from 3 people to 2 people, and streamline the process from 10 steps to 9. step. The results are presented in Tab. 3:

Table 3. Process chartrt after monitoring

Step	Manipulation	Operation	Movement	Inspection	Delay	Storage	Thời gian (s)
1	Remove the power supply from the motor		●				113.2
2	Using a dedicated trolley to load the motor, one person unscrews 4 screws fixing the roller motor	●					124.5
3	Use a hammer and split knife to gently knock the motor away from the fixed chute	●					44.8
4	Take the roller motor out with a dedicated trolley		●				64.5
5	Two people lift the replacement engine onto a dedicated trolley	●					115.1
6	One person adjusts the crank to lift the motor into position, the second aligns the engine flange to the fixed flange via stud bolts.	●					320.8
7	Use a wrench to tighten the 4 flange fixing screws	●					174.5
8	Use your hand to gently rotate the roller to see if there is any friction			●			44
9	Electrical connection and rotation check			●			315
						Total time	**1316.4**

Table 4. Comparison table before and after improvement

No	Method	Ensure saftey	Labor (people)	Time
1	Old method	It is easy to jam hands and feet, back injuries during disassembly	3	1610.1
2	New method: Using a special trolley combined with the planting bolt	Ensure safety during work	2	1316.4
Result of reduction			**1**	**18%**

5. Conclusion

Nowadays, safety in manufacturing is a problem that needs to focus on more and more. This paper presents the utilization of DMAIC methodology, aim in each step-in dismantling and assembly process to find out the unsafety root cause. By applying the Lean-ergonomic combine with the FEMT method, high unsafety risks are removed. As the comparison between before and after applying improving process, it shows that reducing work duration from 1610.2 to 1316.4 (~18%) and reducing 3 employees to 2 employees (~33%) to do the task at roller station. Thereby, not only does it helps workers to work in a safe environment but also work efficiency increases substantially.

Reference

Ahsana Aqilah Ahmad, Amirul Abd Rashid, Freddawati Rashiddy Wong, Muhammad Iqbal. (2017). Journal of Mechanical Engineering. "*Worker Safety Improvement at Paper Pleating Production Line Using Poka-Yoke Concept - A Case Study in Automotive Industry"*. Vol SI 4(5), 183-196.

C. P. Leão, S. Costa (2019). *"Safety Training and Occupational Accidents–Is There a Link?"*.

Dr. U.V. Adinarayana Rao, Santosh Elapanda, Dr. Indukuri Bangar Raju, M. Arun Kumar. (2020) *"Six sigma application in manufacturing special process: A case study on resistance welding"*. International Journal of Advanced Research in Engineering and Technology, 11(7), pp. 568-580.

F. Madariaga. (2013). "*Lean Manufacturing: Exposicion adaptada a la fabricacion repetitiva de familias de productos mediante procesosdiscretos"*.

Federación de Aseguradoras Colombianas Fasecolda, "Riesgos Laborales, estadísticas del ramo – Fasecolda."

Guilherme Tortorella, Johana Cómbita-Niño, Joyce Monsalvo-Buelvas, Lucelys VidalPacheco, Zulmeira Herrera-Fontalvo. (2020) .*"Design of a methodology to incorporate Lean Manufacturing tools in risk management, to reduce work accidents at service companies"*. Procedia Computer Science 177 (2020), 276–283.

Jeroende Mast, JoranLokkerbol. (2012). *"An analysis of the Six Sigma DMAIC method from the perspective of problem solving"*. Int. J. Production Economics 139 (2012) 604–614.

K. Nieuwenhuijsen, D. Bruinvels, and M. Frings-Dresen, *"Psychosocial work environment and stress-related disorders, a systematic review.,"* Occup. Med. (Lond)., vol. 60, no. 4, pp. 277–286, 2010, doi: 10.1093/occmed/kqq081.

L. Briceño (2003), *"Prevención de riesgos ocupacionales en empresas colombianas,"* Rev. Cienc. Salud, vol. 1, no. 1, pp. 31–44..

M. Rajadell Carreras, J. L. Sánchez García. (2010). "*Lean Manufacturing*. La evidencia de una necesidad".

Michael Testani, Sreekanth Ramakrishnan. (2013)."*Ergonomic Risk Reduction to Enhance Lean Transformation"*. Proceedings of the 2013 Industrial and Systems Engineering Research Conference, 989-997.

Monika Smętkowska, Beata Mrugalska. (2018) *"Using Six Sigma DMAIC to Improve the Quality of the Production Process: A Case Study"*.Procedia - Social and Behavioral Sciences 238 (2018), 590 – 596R. Khani, A. Salehi, S. Sajad. (2018). *"Relationship Between Lean Manufacturing and Ergonomics"* Adv. Intell. Syst. Comput., vol. 606, no. October, pp. 1–5, 2018, doi: 10.1007/978-3-319-60474-9.

Saurin, T., Formoso, C., and Cambraia, F., 2006, *"Towards a Common Language between Lean Production and Safety"* Proc. of the 14th Annual Conference of International Group for Lean, July 25-27, Santiago, Chile, 483-495.

Zélio Geraldo dos Santos, Leandro Vieira, Giles Balbinotti. (2015). "*Lean Manufacturing and ergonomic working conditions in the automotive industry"*. Procedia Manufacturing 3 (2015), 5947 – 5954.

VIETNAM ACADEMY OF SOCIAL SCIENCES
SOCIAL SCIENCES PUBLISHING HOUSE
26 Ly Thuong Kiet - Hoan Kiem - Ha Noi
Tel: 024.39719073 - Fax: 024.39719071
Website: http://nxbkhxh.vass.gov.vn
Email: nxbkhxh@gmail.com

Branch in Ho Chi Minh City
57 Suong Nguyet Anh - Ben Thanh Ward - District 1, Ho Chi Minh City
Tel: 028.38394948 - Fax: 028.38394948

PROCEEDINGS OF INTERNATIONAL CONFERENCE ON LOGISTICS AND INDUSTRIAL ENGINEERING 2021

VAN LANG UNIVERSITY - FACULTY OF COMMERCE

LOGISTICS AND SUPLY CHAIN DEPARTMENT

Responsible for Publication:
Assoc. Prof. Dr. PHAM MINH PHUC

Content Editor: NGUYỄN KIM DUNG
Computer Technician: NGUYỄN VIẾT TỊNH
Printing Corrector: NGUYỄN VIẾT TỊNH
Cover Designer: NGUYỄN THÁI NGUYỆT THANH

Number of printed copies: 100, size 19x27 cm, in Công ty TNHH MTV In Đắk Lắk
Address: 45 Nguyễn Tất Thành, phường Tân An, TP. Buôn Ma Thuột, tỉnh Đắk Lắk
Registration number:3578-2021/CXBIPH/3-228/KHXH.
Publication issue number: 325/QĐ-NXB KHXH ngày 6 tháng 12 năm 2021.

ISBN: 978-604-308-723-9.

Published and submitted for catalog record in 12/2021.

MINISTRY OF EDUCATION AND TRAINING
VAN LANG UNIVERSITY

ISSN: 2734-956X

ISBN: 978-604-308-723-9