Lecture Notes in Production Engineering

Chin-Yin Huang · Rob Dekkers ·
Shun Fung Chiu · Daniela Popescu ·
Luis Quezada *Editors*

Intelligent and Transformative Production in Pandemic Times

Proceedings of the 26th International Conference on Production Research

Springer

Lecture Notes in Production Engineering

Lecture Notes in Production Engineering (LNPE) is a new book series that reports the latest research and developments in Production Engineering, comprising:

- Biomanufacturing
- Control and Management of Processes
- Cutting and Forming
- Design
- Life Cycle Engineering
- Machines and Systems
- Optimization
- Precision Engineering and Metrology
- Surfaces

LNPE publishes authored conference proceedings, contributed volumes and authored monographs that present cutting-edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of excellence. Also considered for publication are lecture notes and other related material of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of production engineering. The target audience of LNPE consists of advanced level students, researchers, as well as industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNPE will be distributed through Springer's print and electronic publishing channels. To submit a proposal or request further information please contact Anthony Doyle, Executive Editor, Springer (anthony.doyle@springer.com).

Chin-Yin Huang · Rob Dekkers · Shun Fung Chiu ·
Daniela Popescu · Luis Quezada
Editors

Intelligent and Transformative Production in Pandemic Times

Proceedings of the 26th International Conference on Production Research

Springer

Editors
Chin-Yin Huang
Department of Industrial Engineering
and Enterprise Information
Tunghai University
Taichung, Taiwan

Shun Fung Chiu
College of Engineering
De La Salle University
Manila, Philippines

Luis Quezada
Department of Industrial Engineering
Universidad de Santiago de Chile
Santiago, Chile

Rob Dekkers
Adam Smith Business School
University of Glasgow
Glasgow, UK

Daniela Popescu
Universitatea Tehnica din Cluj-Napoca
Cluj-Napoca, Romania

ISSN 2194-0525 ISSN 2194-0533 (electronic)
Lecture Notes in Production Engineering
ISBN 978-3-031-18640-0 ISBN 978-3-031-18641-7 (eBook)
https://doi.org/10.1007/978-3-031-18641-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Due to the influence of the COVID-19 pandemic, the 26th International Conference on Production Research (ICPR) online was hosted by Tunghai University, Taiwan, in July 2021. It was the first time that ICPR was ever held in Taiwan. The decision on the theme of the conference "Intelligent and Transformative Production in Pandemic Times" was a collaborative work by the board members of International Foundation for Production Research (IFPR) and Asia Pacific Region of IFPR.

The keynote speakers of the conference, Prof. Shimon Nof, Prof. Oliver Riedel, Prof. Alexandre Dolgui, Dr. Kenichi Funaki and Prof. Morris Fan, presented their cutting-edge and inspiring research and ambition toward the future of production research. Those speeches highlighted the conference. The Doctoral and Early Career Researcher Workshop organized by Prof. Rob Dekkers prior to the conference was once again a successful and memorable one.

This proceedings book of the 26th ICPR selected 82 papers from the conference. More than 250 scholars/students participated in the conference. Papers to be included in this book are within the following twelve subjects:

Part One: Impact of the Pandemic on Industry and Production
Part Two: Digital and Cyber Manufacturing and Services
Part Three: Manufacturing: System and Automation
Part Four: Internet of Things, Data Analytics and Smart Manufacturing
Part Five: Supply Chain Management
Part Six: Industrial Engineering and Operations Research
Part Seven: Logistics Engineering and Management
Part Eight: Sustainable Production
Part Nine: New Product Development and Innovation Management
Part Ten: Quality Engineering and Management
Part Eleven: Human Factors Engineering
Part Twelve: Healthcare Management

The papers show the courage and faith of the authors across the world by demonstrating their cutting-edge knowledge and wisdom to face the near bright future. The five editors from three continents would like to show our appreciation and respect to the authors' contributions and in the same time to recommend their precious work to the readers worldwide. May this book become a courageous marker of production research at the very difficult time of the human world during the pandemic!

Finally, we would like to thank every contributor/participant/reviewer/session chair of the conference. Thanks go to the Ministry of Science and Technology, Taiwan, for the funding support, to Tunghai University for the facility support, to the Department of Industrial Engineering and Enterprise Information at Tunghai University for the assistance from the staffs and students, to JIMA of Japan, KIIE of Korea, PIIE of Philippines, CIIE of Taiwan, IFPR-AEM, IFPR-AR and IFPR-APR for the promotions of the conference. We thank Lecture Notes in Production Engineering, Springer, for publishing this book.

Taichung, Taiwan — Chin-Yin Huang
Glasgow, UK — Rob Dekkers
Manila, Philippines — Shun Fung Chiu
Cluj-Napoca, Romania — Daniela Popescu
Santiago, Chile — Luis Quezada

Contents

Digital and Cyber Manufacturing and Services

Manufacturing: System and Automation

Impact of the Pandemic on Industry and Production

Development of Training System for Appearance Inspection Using Motion Capture and Large Size Display

Ryosuke Nakajima and Jyunya Fujimoto

Abstract To assure appearance quality of industrial products, appearance inspection by human vision is implemented in many manufacturing industries. In appearance inspection by human vision, small visual defects such as dirt, scratches, surface dents, and unevenness of the coating color are inspected. In actual defects in appearance inspection processes, the visibility differs depending on the positional relationship between the inspector and the defect. Furthermore, visibility varies depending on the type and characteristic of defects. Therefore, to comprehensively inspect these defects, it is required to change the positional relationship between the inspector and defects by appropriate body movements of the inspector. In particular, in case of product size is relatively large such as a body, glass of an automobile, there is strongly required the body movement of the inspector to realize positional relationship in which defects can be detected. However, no studies have been conducted on appropriate body movements based on the defect positional relationship between the inspector and defects, in actual appearance inspection processes, work training is only carried out by on the job training based on experience of skilled inspectors. Therefore, this study focused on appearance inspection that required appropriate body movements of inspectors during the inspection, and also developed a work training system for appearance inspection using motion capture and large size display. Moreover, to show how to utilize the developed training system, the experiment using the general work procedure manual and the work procedure manual that teaches detailed body movements according to size of products and height of an inspector are designed as experimental factors, and their effects on defect detection are evaluated with 12 subjects. As a result, it is quantitatively obtained that the latter work procedure manual has the effect of improving the defect detection rate about 5.9%. In particular, it is also obtained to be effective in case of detecting small luminance contrast and detectable range of defects. From the above, it is considered that the

R. Nakajima (✉)
The University of Electro-Communications, Tokyo, Japan
e-mail: nakajima@uec.ac.jp

J. Fujimoto
Seikei University, Tokyo, Japan
e-mail: us153117@cc.seikei.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_1

work training methodology including body movements is effective, and it is shown that the developed training system may contribute to work design and work training in the actual appearance inspection processes.

Keywords Appearance inspection · Visual inspection · Work training · Defect detection · Work design

1 Introduction

To supply high-quality products into the market, manufacturing industries provide attention to product inspection. Two types of inspections exist which are functional inspection and appearance inspection. The former inspects the effectiveness of a product, and the latter, small visual defects, such as scratches, surface dents, and unevenness of the coating color. The automation of functional inspection has advanced because it is easy to determine whether a product works or not [1, 2]. However, it is not as simple to establish standards to determine whether the appearance of a product is defective. First, many different types of defects exist. Second, the categorization of a product as non-defective or defective is affected by the size and depth of the defect. Third, some products have recently become smaller and more complex. Finally, the production has shifted to high-mix and low-volume production. It is thus difficult to develop technologies that can capture small defects and to create algorithms that can identify multiple types of defects with high precision. Therefore, appearance inspection still strongly depends on human vision [3–8]. As appearance inspection is performed by humans, inspection efficiency (inspection time) and inspection accuracy (defect detection rate) differ among different inspectors. This is a common problem in many manufacturing industries.

Moreover, in actual defects in appearance inspection processes, the visibility differs depending on the positional relationship between the inspector and the defect [9]. Furthermore, visibility varies depending on the type and characteristic of defects [10]. Therefore, to comprehensively inspect these defects, it is required to change the positional relationship between the inspector and defects by appropriate body movements of the inspector. In particular, in case of product size is relatively large such as a body, glass of an automobile, there is strongly required the body movement of the inspector to realize positional relationship in which defects can be detected. There are some studies on work training on how to move the viewpoint for relatively small products such as printed circuit boards [11]. There are also some studies on defect visibility assessment [12, 13] and work design methodologies [14–16]. However, no studies have been conducted on appropriate body movements based on the defect positional relationship between the inspector and defects, in actual appearance inspection processes, work training is only carried out by on the job training based on experience of skilled inspectors.

Therefore, this study focused on appearance inspection that required appropriate body movements of inspectors during the inspection, and also developed a work training system for appearance inspection using motion capture and large size display. Moreover, using the training system, this study also considered their effects of work procedure manuals on defect detection experimentally.

2 Development of Training System

2.1 Concept of Training System

For work design and training of appearance inspection of large size products, the concept of detectable range is necessary. This is the range in which defects can be detected based on positional relationship between a defect and head of an inspector, and is represented by a square centered on the defects. Based on the concept of detectable range, this study considered to determine the body movements to reliably detect defects that exist everywhere. Specifically, to consider the work design method for large products, characteristics of defects, height of inspectors, size of a product, and detectable range of a defect are determined as input information. Then, the height at which the product is installed is obtained from the height of the inspector, and the appropriate body movements are designed as output information.

That is, a condition in which a defect can be detected from the positional relationship between the inspector and the defect is defined as a detection condition, the body movement data of the inspector is measured in real time using motion capture, developed the training system that displays defects on the large size display only when the values and the preset detection conditions are identified. The features of the training system are that off the job training is possible by defining only the detection conditions for the target defects in the actual appearance inspection processes, practical work training is possible for new inspectors without preparing actual products, and it is possible to give immediate feedback to the inspectors by totaling the work results in real time.

2.2 Overview of the Developed Training System

Based on the concept in Sect. 2.1, a training system was developed using a motion capture (Microsoft Inc., Xbox One Kinect), a large display (Sony Inc., KJ-49X8500F), and a computer (Panasonic inc., CF-SZ5). Figure 1 shows the overall picture of the training system. First, the position of the inspector's head is measured by motion capture from behind the inspector. It is then determined based on a predefined detectable range. Then, when the positional relationship between the inspector's head and the detectable range is identified, the defect is displayed on the large display. If not identified, the defect is not displayed on the large size display.

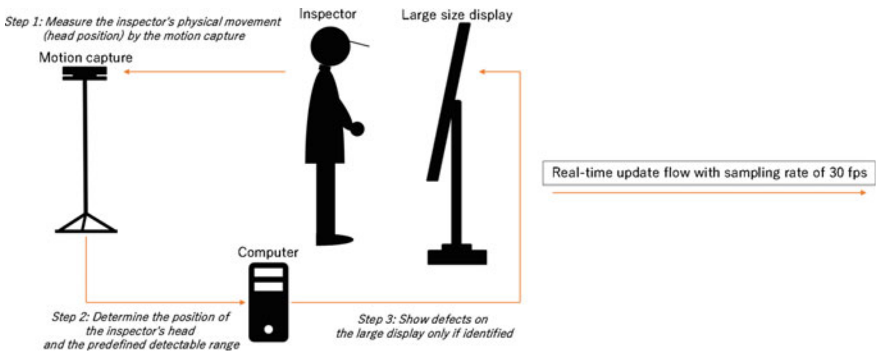


Fig. 1 Conceptual diagram of the training system

3 Experimental Design

3.1 Experimental Tasks

The experimental subjects are tasked with visually inspecting a model that is displayed on the large size display. A model with height 680 mm and width 1200 mm is used. The background color of the inspection model is set to an achromatic color. If no defect is detected, the subject presses the SPACE KEY of the finger mouse, and the next inspection model is displayed. If a defect is detected, the subject presses the ENTER KEY of the finger mouse.

3.2 Experimental Factors

3.2.1 Work Procedure Manual

Two levels of work procedure manuals are used: the general one and the devised one in Sect. 2.1. There is no mention of body movements in the general one.

3.2.2 Luminance Contrast of Defect

The defect characteristics are defined by the luminance contrast between the inspection model and the defect. The defect size is set at 2.5 mm. The luminance contrast of each defect is one of the three following levels: 0.15, 0.20, and 0.25. These defects are determined by assuming the standard of the appearance inspection. Also, the inspection model is divided into 12 parts (4 horizontally, 3 vertically), and the defect is located at the center of either one of those parts.

3.2.3 Detectable Range of Defect

The detectable range of defect is one of the three following levels: 120 mm, 180 mm, and 250 mm. These defects are determined by assuming the standard of the appearance inspection.

3.3 Experimental Procedure

Twelve subjects (two women and ten men), aged between 21 and 24 years with a corrected eyesight score (decimal visual acuity) higher than 1.0, were employed in the experiment. To familiarize the subjects with the experiment, an overview was provided and the experiment procedure was explained. In addition, the subjects were requested to perform some preliminary experiments. In the experiment, the task was to inspect 216 inspection models.

The experimental room temperature was set between 18 and 24 °C, and the humidity was set between 40 and 60%. Because the luminance of the inspection model and the defects were affected by external and internal light (such as fluorescent lighting), the experiment was implemented in a dark room. A written statement of the purpose and contents of the experiment was provided to the subjects and informed consent was obtained from all subjects.

4 Experimental Results

4.1 Individual Characteristics of Subjects

Using the defect detection rate, the effect of the training method is examined. Owing to the possibility that the individuality of the subjects might influence the results, the uniformity of the results for all subjects is verified.

The defect detection rate of each subject is shown in Fig. 2. As a result of the Smirnov–Grubbs test, no outlier values are observed in the defect detection rates of any of the subjects. Therefore, all the data from twelve subjects are used.

4.2 Effect of Work Procedure Manual on Defect Detection

To analyze the effect of work procedure manuals on defect detection, three-way analysis of variance (ANOVA) is executed with the work procedure manual (2), luminance contrast of defect (3), and detectable range of defect (3) as factors. The ANOVA table is shown in Table 1. As a result, a significant difference of 1% is

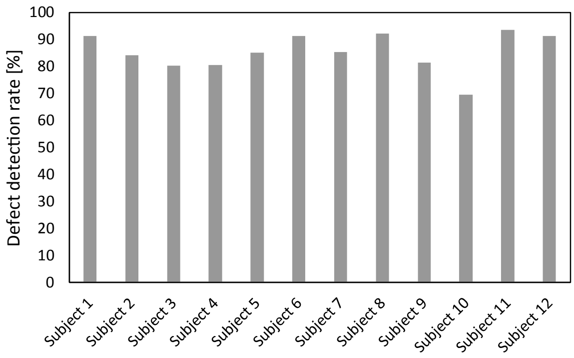


Fig. 2 Defect detection rate for each subject

observed for the main effect of the luminance contrast of defect, the detectable range of defect. Furthermore, a significant difference of 5% is observed for the main effect of the work procedure manual, and mutual interactions between the work procedure manual and the luminance contrast of defect, the work procedure manual and the detectable range of defect. The effect of these experimental factors on defect detection are shown in Fig. 3, the relationship between each experimental factors are shown in Fig. 4.

From this results, it is clarified that the devised work procedure manual has the effect of improving the defect detection rate about 5.9% as compared with the general work procedure manual. In particular, it is shown to be effective in case of detecting small luminance contrast and detectable range of defects. In the future, we will analyze the differences between these work procedure manuals more in detail using the motion capture data of subjects.

Based on these findings, by using the training system developed in this study, it is possible to quantitatively examine the effects of work procedure manuals and/or differences in skills between workers on defect detection. It is considered to contribute to work design and work training of the actual appearance inspection process.

5 Conclusion

This study focused on appearance inspection that required appropriate body movements of inspectors during the inspection, and also developed a work training system for appearance inspection using motion capture and large size display. Specifically, a condition in which a defect can be detected from the positional relationship between the inspector and the defect is defined as a detection condition, the body movement data of the inspector is measured in real time using motion capture, developed the training system that displays defects on the large size display only when the values and the preset detection conditions are identified. The features of the developed training system are that off the job training is possible by defining only the detection

Table 1 ANOVA table

Factor	Sum of squares	Degrees of freedom	Mean square	F-value	Significant difference
Subject (S)	1.01	11	0.09		
Work procedures manual (A)	0.19	1	0.19	7.75	*
S × A	0.27	11	0.02		
Luminance contrast of defect (B)	0.69	2	0.35	44.12	**
S × B	0.17	22	0.01		
Detectable range of defect(C)	0.99	2	0.49	46.35	**
S × C	0.23	22	0.01		
A × B	0.07	2	0.03	4.80	*
S × A × B	0.16	22	0.01		
A × C	0.04	2	0.02	5.26	*
S × A × C	0.09	22	0.00		
B × C	0.07	4	0.02	1.72	
S × B × C	0.45	44	0.01		
A × B × C	0.05	4	0.01	3.13	*
S × A × B × C	0.19	44	0.00		
Total	4.68	215			

conditions for the target defects in the actual appearance inspection processes, practical work training is possible for new inspectors without preparing actual products, and it is possible to give immediate feedback to the inspectors by totaling the work results in real time.

Moreover, using the developed training system, a subject experiment was conducted to compare a general work procedure manual with a work procedure manual that also teaches detailed body movements. As a result, it was found that the latter work procedure manual has the effect of improving the defect detection rate. In particular, it has been shown to be effective in detecting defects such as luminance contrast and a small detectable range. From the above, it is considered that the work training methodology including body movements is effective, and it is shown that the developed training system may contribute to work design and work training in the actual appearance inspection processes.

Future studies may consider applying these findings to the manufacturing industry and examining the results.

Fig. 3 Average value of experimental factors. **a** Work procedure manual. **b** Luminance contrast of defect. **c** Detectable range of defect

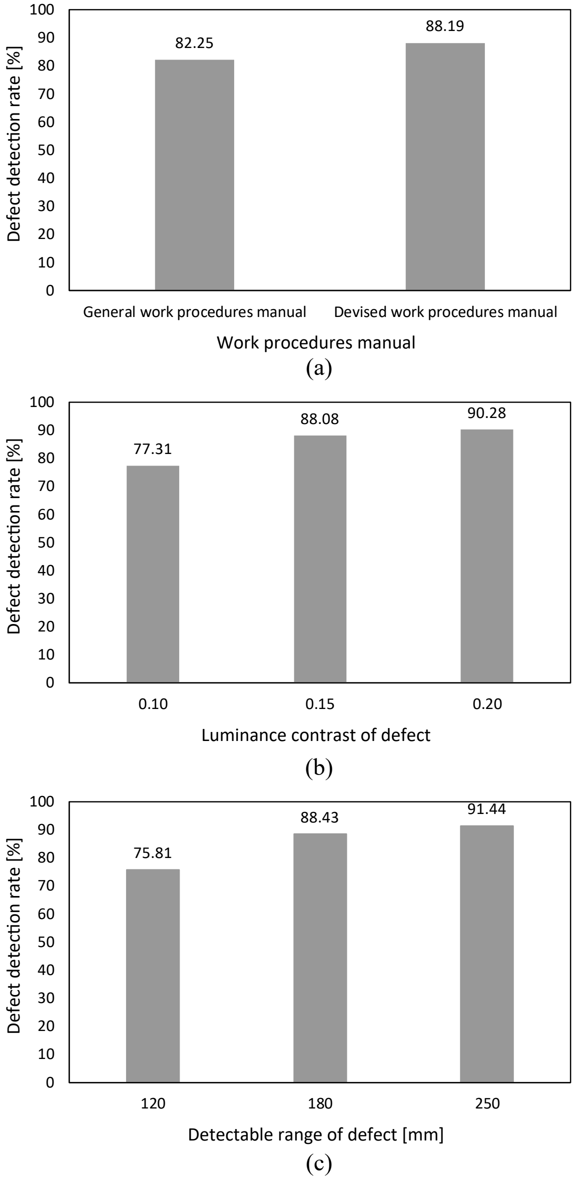
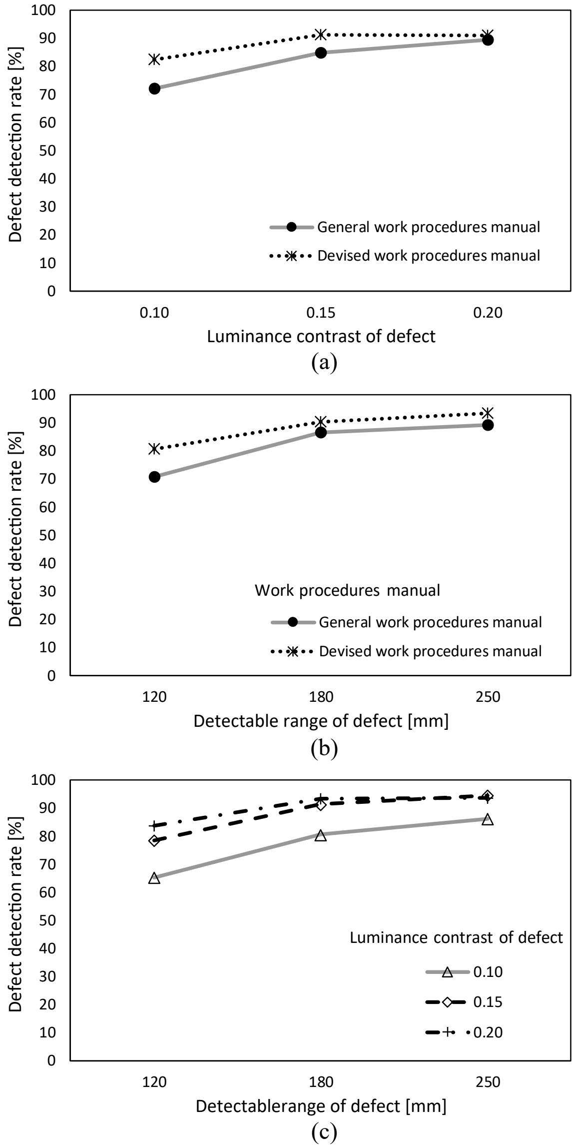


Fig. 4 Interaction between experimental factors. **a** Relationship between (A) * (B). **b** Relationship between (A) * (C). **c** Relationship between (B) * (C)



Acknowledgements This study was partly supported by JSPS Grant-in-Aid for Young Scientists (19K15237).

References

1. Aoki, K.: Review on in-process inspection. J. Japan. Soc. Non-Destr. Inspection **63**(8), 433–435 (2014)
2. Aiyama, H.: Review on magnetic particle, penetrant and visual testing. J. Japan. Soc. Non-Destr. Inspection **63**(8), 384–390 (2014)
3. Nickles, G.M., Melloy, B.J., Gramopadhye, A.K.: A comparison of three levels of training designed to promote systematic search behavior in visual inspection. Int. J. Ind. Ergon. **32**, 331–339 (2003)
4. Yeow, P.H., Sen, R.N.: Ergonomics improvements of the visual inspection process in a printed circuit assembly factory. Int. J. Occup. Saf. Ergon. **10**(4), 369–385 (2004)
5. Lee, F.C., Chan, A.H.: Effects of magnification methods and magnifier shapes on visual inspection. Appl. Ergon. **40**(3), 410–418 (2009)
6. Chang, J.J., Hwang, S.L., Wen, C.H.: The development of a training expert system for TFT-LCD defects inspection. Int. J. Ind. Eng.Theory Appl. Pract. **16**(1), 41–50 (2009)
7. Ishii, A.: Now Is the time for reviewing visual inspection. J. Japan. Soc Precis. Eng. **84**(12), 963–966 (2018)
8. Johnson, T.L., Fletcher, S.R., Baker, W., Charles, R.L.: How and why we need to capture tacit knowledge in manufacturing: case studies of visual inspection. Appl. Ergon. **74**(1), 1–9 (2019)
9. Nakajima, R., Shida, K., Matsumoto, T.: A study on the effect of irradiation angle of light on defect detection in visual inspection. Int. J. Ind. Eng. Theor. Appl. Pract. **22**(1), 525–537 (2015)
10. Nakajima, R., Nakamura, R., Hida, T., Matsumoto, T.: Effect of bright and shade, and luminance difference of defect on defect detection in appearance inspection utilizing peripheral vision. Int. J. Ind. Ergono. **82**(103086) (2021). https://doi.org/10.1016/j.ergon.2021.103086
11. Nakajima, R., Inagaki, I., Matsumoto, T.: Development of a training system for visual inspection method utilizing peripheral vision. J. Japan. Ind. Manag. Assoc. **66**(3), 267–276 (2015). https://doi.org/10.11221/jima.66.267
12. Nakajima, R., Asano, Y., Hida, T., Matsumoto, T.: The relationship between dirt levels of inspection surface and defect detection in visual inspection utilizing peripheral vision. Ind. Eng. Manag. Syst. **17**(1), 102–112 (2018). https://doi.org/10.7232/iems.2018.17.1.102
13. Nakajima, R., Yamamoto, R., Hida, T., Matsumoto, T.: A study on the effect of defect shape on defect detection in visual inspection. In: 25th International Conference on Production Research 2019, Paper ID 234 (2019)
14. Nakajima, R., Tanida, K., Hida, T., Matsumoto, T.: A study on the relationship be-tween defect characteristics and conspicuity field in visual inspection. Japan. J. Ergon. **51**(5), 333–342 (2015). https://doi.org/10.5100/jje.54.153
15. Nakajima, R., Uchino, U., Hida, T., Matsumoto, T.: A study on the effect of luminance of light on defect detection in visual inspection utilizing peripheral vision. Japan. J. Ergon. **54**(4), 153–163 (2018). https://doi.org/10.5100/jje.54.153
16. Nakajima, R., Suzuki, Y., Hida, T., Matsumoto, T.: A study on the effect of luminance of inspection surface on defect detection in visual inspection utilizing peripheral vision. Japan. J. Ergon. **54**(1), 14–23 (2018). https://doi.org/10.5100/jje.54.14

Worker Organization System for Collaborative Crowdsourcing

Ryota Yamamoto and Kazushi Okamoto

Abstract Crowdsouring with cooperative collaboration among experts should be performed by teams that consider compatibility among crowd workers. In this study, we develop a system that can automatically form an organization to efficiently perform complex and large-scale projects in crowdsourcing. The system targets projects that can be divided into multiple tasks, and the output organization consists of subgroups and subleaders that correspond to the tasks. The compatibility among crowd workers is calculated by weighted social network generated with the past experience of collaboration. Furthermore, we develop an algorithm based on the greedy method to search teams with compatibility optimization under the constraints of budget and skills using the social network. In the experiment, we collected 169,627 users on 33,983 repositories from GitHub, and determined the social network and skills of the workers by the programming languages, topics, and user contributions from the repositories. The characteristics of the developed system are observed by the simulation experiments with a virtual project with skills and budget requirements. In particular, the comparison of teams formed with minimum budget suggests that the proposed team formation algorithm achieves a 96% higher compatibility among crowd workers than the algorithms that do not consider compatibility. Furthermore, the proposed algorithm was compared with the algorithm of the full search in related work, and it was found that full search algorithm enhances the team compatibility when a single team is formed under strict budget constraints.

Keywords Crowdsourcing · Social network · Combinatorial optimization · Team organization

R. Yamamoto · K. Okamoto (✉)
The University of Electro-Communications, Tokyo, Japan
e-mail: kazushi@uec.ac.jp

R. Yamamoto
e-mail: r.yamamoto@uec.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_2

1 Introduction

Crowdsourcing is a practice/sourcing model in which individuals or organizations entrust an unspecified number of people, hired through the Internet, to perform specific tasks. It has enabled human resource management depending upon the scale and property of tasks [1]. In particular, collaborative crowdsourcing has been attracting considerable attention because it forms teams among crowd workers with professional knowledge and skills to accomplish complex tasks such as software development. Collaborative team work is expected to have synergistic effects such as increased work motivation [2, 3]. Conversely, improper communication among crowd workers can cause shortcomings in collaborative works. These include lack of information sharing and discord in the team. To form teams that collaborate efficiently in crowdsourcing, it is necessary to consider the compatibility that represents the strength of real-world social connections between workers [4].

The number of team members is proportional to the size of the requested task. This might cause compatibility issues in the team owing to the increase in the number of people required for communication. Therefore, when a large-scale task (project) is divided into several subtasks, it is desirable to form subgroups for each subtask to construct an organization. For example, a software development project can be divided into tasks such as front-end, back-end, and mobile development, and teams can be assigned to perform each task. Hence, forming teams based on the organizational structure can decrease the number of members in each team and clarify the roles of workers as well as the chain of command.

In this study, we propose a system that constructs an online organization to perform a complex and large-scale project through crowd workers with specialized abilities at a limited budget. We solve a combinatorial optimization problem maximizing the compatibility within the organization using employment costs and skill requirements as constraints, and use a social network (SN) of crowd workers that represents social relationships in the real world. In addition, we observe the characteristics of the proposed system through experiments using the data collected from GitHub.

2 Related Works

Valentine et al. proposed a system for constructing an organization in which hierarchy, roles, and a chain of command are clarified to perform complex and open-ended tasks [5]. In this system, compatibility among workers is not considered because workers are hired through open recruitment. However, the workers can decide their upcoming actions and their points of communication through the clarification of roles and a chain of command of workers. This enables collaborative work with workers having no common experience.

In the SN used in this study, workers and relationships are represented as nodes and edges, respectively. The edge is dictated by past work-related connections such as colleagues, superiors, etc. The weight of the edges denotes the collaboration cost. The higher the compatibility is, the smaller the collaboration cost becomes. Lappas et al. formulated the problem of forming a team with optimized compatibility, constrained by the professional ability of the entire team [6]. Gajewar et al. proposed a new measure of compatibility called density, which is defined as the sum of the weights of the edges with respect to the number of nodes in the team [7]. Anagnostopoulos et al. considered the number of tasks a worker can perform [8]. Wang et al. designed an incentive system which prevents crowd workers from falsifying their own profiles [9]. However, the algorithms proposed in these studies form a single team for a single task and cannot deal with the problem proposed in this study: formation and organization of multiple subgroups from the same worker pool.

Rahman et al. proposed an algorithm that forms a team with compatibility optimized and divides it into several subgroups to limit the size of the team [10]. In the algorithm proposed by Rahman et al., there are two algorithms: one for determining the set of workers to be employed and the other for dividing the team. The former algorithm forms a team $\mathcal{G}$ to minimize a measure of compatibility $DiaDist(\mathcal{G})$ with the skill level and budget requirements as constraints,

$$DiaDist(\mathcal{G}) = \max_{u_i, u_j \in \mathcal{G}} dist\left(u_i, u_j\right) \tag{1}$$

where $dist\left(u_i, u_j\right)$ represents the compatibility between u_i and u_j and is predetermined for all worker combinations. For a given worker, the algorithm implements a function that can perform a full search for a set of workers whose *dist* with that worker is less than or equal to a threshold α. Furthermore, the process is repeated for various workers until a team that satisfies the constraint condition is found. In this way, it can form a team with $DiaDist(\mathcal{G}) < 2\alpha$. The algorithm performs a binary search for α to minimize $DiaDist(\mathcal{G})$. The latter algorithm divides $\mathcal{G}$ to subgroups $\{G_1, G_2, \ldots, G_x\}$ using the number of team members as constraints. For a given a set of workers $\{u_1, u_2, \ldots, u_x\}$ corresponding to each subgroup, this algorithm repeats a function that allocates other workers to the subgroups minimizing $\sum_{i=1}^{x} k_i \times dist(u_i, \cup_{j \neq i} G_j) + \sum_{i<j} k_i \times k_j \times dist(u_i, u_j)$ for all combinations of $\{u_1, u_2, \ldots, u_x\}$, where k_i represents the size of G_i. In this study, unlike that undertaken by Rahman et al., there are multiple tasks, and teams are formed for each task. Therefore, it is difficult to divide the teams after determining the set of workers to be employed as mentioned in the Rahman et al.'s algorithm.

3 Proposed Worker Organization System

In this study, we propose a system that automatically constructs an organization that satisfies budget and skill requirements and optimizes the compatibility of the system proposed by Valentin el al. In the proposed system, a project leader is assigned and the project is divided into multiple tasks. The project leader is assumed to be a requester, and is not decided by the proposed system. Since each task has a set of skill requirements, it is difficult to divide the team after determining the set of workers to be employed as mentioned in the study by Rahman et al. We form subgroups that fulfill the requirements of expertise and budget for each divided task, and assign a leader to each team.

In the proposed system, there is a set of workers $U = \{u_1, u_2, \ldots, u_m\}$. The project is defined by a set of tasks $T = \{\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_n\}$ and budget b. A task $\boldsymbol{t} \in T$ is represented by a l-dimensional required skill level vector $[t_1, t_2, \ldots, t_l]$. Each worker u has a skill level vector $[s_1, s_2, \ldots, s_l]$ that denote level of skills which a worker u possesses and employment cost c. Collaboration cost $d(u, v)$ is estimated by SN based on past work-related connections ($d(u, u) = 0$).

We first solve the problem that forms the subgroup $\mathcal{G} = \{G_1, G_2, \ldots, G_n\}$ for a task v. Each subgroup G_j should fulfill cost requirements $\sum_{j=1}^{n} \sum_{u \in G_j} c_u \leq b$ and skill requirements $\sum_{u \in G_j} s_k^u \geq t_k^j \forall k \in [1, 2, \ldots, l]$. Under these conditions, we find the Pareto optimal solution for the average of maximum distance on team $f_d(\mathcal{G}) = \sum_{j=1}^{n} \max_{u,v \in G_j} d(u, v)/n$ and the sum of distances on team $f_s(\mathcal{G}) = \sum_{j=1}^{n} \sum_{u,v \in G_j} d(u, v)$. We solve this combinatorial optimization problem by the greedy algorithm. Let

$$e_1(u, j) = \frac{\sum_{k=1}^{l} \min(\max(t_k^j - \sum_{v \in G_j} s_k^v, 0), s_k^u)}{c_u} \tag{2}$$

be the utility that occurs when u is assigned to G_j.

Algorithm 1: *MinDia* allocates root worker u_{r_j} to each G_j at the beginning. u_{r_j} is the worker with the highest utility $e_1(u, j)$ among workers who tend to have low collaboration cost with others. The G_j is formed from the workers in the social network around each root worker. In this process, threshold τ is set and G_j is formed from workers whose distance from u_{r_j} is less than or equal to τ. The smallest threshold τ' is determined through binary search and minimizing $f_d(\mathcal{G})$.

Algorithm 1: *MinDia*

Input: U, T

1: for each $j \in [1, 2, \ldots, n]$

2: $\quad u_{r_j} \leftarrow \underset{u \in U}{\text{argmax}}\ e_1(u, j)$ such that magnitude of $\sum_{v \in U} d(u, v)/(n-1)$ is lower 1% in U

3: $\tau_L \leftarrow 0, \tau_U \leftarrow \max_{u,v \in U} d(u, v)$

4: repeat any number of loops

5: $\tau \leftarrow (\tau_L + \tau_U)/2, \mathcal{G} \leftarrow FormTeam\big(\tau, \{u_{r_1}, u_{r_2}, \ldots, u_{r_n}\}\big)$

6: if $\mathcal{G} \neq \emptyset$ then $\tau_U \leftarrow \tau, \mathcal{G}' \leftarrow \mathcal{G}$ else $\tau_L \leftarrow \tau$

7: return $\mathcal{G}'$

Algorithm 2: *FormTeam*

Input: $U, T, b, \tau, \{u_{r_1}, u_{r_2}, \ldots, u_{r_n}\}$

repeat

$u', j' \leftarrow \underset{u \in U \setminus (G_1 \cup G_2 \cup \ldots \cup G_n), j \in [1,\ldots,n]}{\text{argmax}} e(u, j)$ such that $dist(u, u_{r_j}) \leq \tau$ and

$\sum_{j=1}^{n} \sum_{v \in G_j} c_u + c_v \leq b$

if u', j' are found and $e_1(u', j') > 0$ then

$G_{j'} \leftarrow G_{j'} \cup \{u'\}$

while $\exists u'' \leftarrow \underset{u \in G_{j'}}{\text{argmax}}\, c_u$ such that $\sum_{k=1}^{l} \min(\max(t_k^j - \sum_{v \in G_{j'} \setminus \{u\}} s_k^v, 0), s_k^u) = 0$

$G_{j'} \leftarrow G_{j'} \setminus \{u''\}$

else if all G_j fulfill cost and skill requirements

return $\{G_1, G_2, \ldots, G_n\}$

else

return $\emptyset$

Algorithm 3: *ChangeWorker*

Input: $\mathcal{G}', U, T, b, \{u_{r_1}, u_{r_2}, \ldots, u_{r_n}\}$

repeat

$j', u'_{out1}, u'_{out2}, u'_{in} \leftarrow \underset{j \in [1,\ldots,n], u_{out1}, u_{out2} \in G_j, u_{in} \in U \setminus \mathcal{G}}{\text{argmax}} e_2(j, \{u_{out1}, u_{out2}\}, u_{in})$

such that $d(u_{in}, u_{r_j}) \leq \tau$ and $G_j \setminus \{u_{out1}, u_{out2}\} \cup u_{in}$ fulfill cost and skill requirements

if $j', u'_{out1}, u'_{out2}, u'_{in}$ are found and $e_2(j', \{u'_{out1}, u'_{out2}\}, u'_{in}) > 0$ then

$G_{j'} \leftarrow G_{j'} \setminus \{u'_{out1}, u'_{out2}\} \cup \{u'_{in}\}$

else

break

repeat same process with one-to-one exchange

return $\{G_1, G_2, \ldots, G_n\}$

Subroutine algorithm 2: *FormTeam* forms each G_j based on the greedy method. The algorithm finds the combination of u and G_j, resulting in the highest utility and fulfills $d(u, u_{r_j}) \leq \tau$; furthermore, it performs allocation for the combination. This process is continued until the constraints are satisfied. If the total cost of hiring all subgroups exceeds the budget without fulfilling the skill requirements, or if there are no more workers who fulfill the requirements by τ, the process is stopped. An extra worker can be created in a loop. For example, when skill requirement vector is (2, 2), and sum of skill level vector of a subgroup is (4, 2), a worker in the subgroup whose skill vector is (2, 0) is redundant. Therefore, if such an extra worker is found after a new worker is added to G_j in a loop, the worker will be removed from G_j.

If there are more than one, the worker with highest cost is removed. This process is repeated until the extra workers are excluded.

The greedy method considers workers with the lowest hiring cost for their skill level, even when the budget is sufficient. Therefore, the total employment cost of the entire team might be much lower than that of the budget allocated. The algorithm 3: *ChangeWorker* exchanges workers for teams formed by the greedy method to ensure that the $f_s(G)$ is as small as allowed by the budget. *ChangeWorker* finds a combination of exchange workers minimizing $f_s(G)$ and fulfilling the requirement by τ'. Here, when the set of workers R is released and a new worker u is added in G_j, the utility function $e_2(j, R, u)$ is defined

$$e_2(j, R, u) = \frac{f_s\big((G_j \backslash R) \cup \{u\}\big) - f_s\big(G_j\big)}{c_u}. \tag{3}$$

The algorithm first exchanges two workers in the team for one external worker. When workers are unavailable for two-to-one exchange, it performs one-to-one exchanges.

Finally, subleaders $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ while minimize $\sum_{a_i, a_j \in \mathcal{A}} d\big(a_i, a_j\big)$ is determined to each G_j by full search about all combination of $\mathcal{A}$.

4 Evaluation Experiment

In the experiment, we collected data from 33,983 repositories and 169,627 users from GitHub to generate worker profiles, SN, and virtual project. The SN is determined by the relationship among the contributors. A collaboration cost between a pair of workers is calculated from the sum of the smaller contributions in repositories, in which both workers contribute. In this case, we consider the reciprocal of the sum. The direct product set of the topics and the programming languages used on the repository is defined as the skill, and the skill level is calculated from the ratio of the file size of each language and the user's contribution. As for the employment cost, we estimate the hourly wage based on the skill level, the average annual income by programming languages [11], and the global average working hours [12]. We assume that the cost distribution of workers is an exponential distribution. It is found that each skill level distribution approximately follows an exponential distribution. Therefore, we represent hourly wage of a worker by multiplying the skill level of the worker with the ratio of the average cost to the average skill level. However, in practice, hourly wage differs based on languages. We calculate hourly wage for each language, and use the average of the top 3 as hourly wage.

Table 1 Comparison to the baseline algorithm (Ex. 1)

Skill level	Budget	Baseline		Proposed method	
		$f_d(\mathcal{G})$	$f_s(\mathcal{G})$	$f_d(\mathcal{G})$	$f_s(\mathcal{G})$
20	680	23.7	9.06×10^3	20	680
25	840	26.9	12.4×10^3	25	840
30	995	25.9	17.6×10^3	30	995
35	1165	28.2	26.9×10^3	35	1165
40	1325	28.4	32.6×10^3	40	1325
Average		26.6	19.7×10^3	1.16	26.6

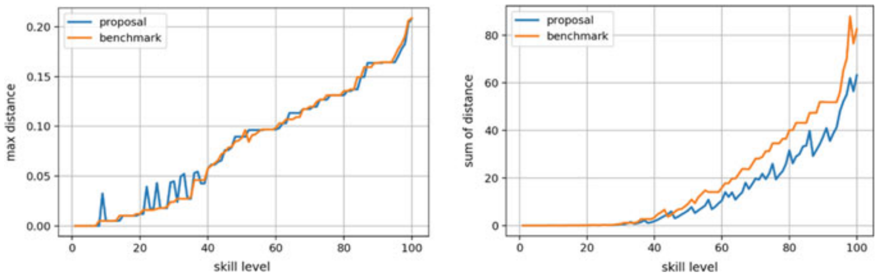
4.1 Experiment 1: Comparison with the Greedy Method Without Taking into Account Compatibility

We compared the proposed algorithm with the greedy algorithm, which does not consider compatibility, i.e., the threshold τ is infinite. There are three tasks of the virtual project used in this experiment. The first task is (“nodejs”, “JavaScript”), (“nodejs”, “ HTML/CSS”), (“nodejs”, “SQL”), the second is (“android “, “Java”), (“android”, “Kotlin”), and the third is (“ios”, “Swift”), (“ios”, “Objective-C”). The system forms teams for all combinations of skill levels in the range of {20, 25, 30, 35, 40} and budgets {650, 655, …, 2000}. For each skill level, we compared the teams at the smallest budget for team formation.

The result of experiments is summarized in Table 1. Comparing the averages of each index of the two algorithms, we can see that the proposed algorithm improves the maximum distance on team by approximately 95.64%, and the sum of distances on team by about 99.48% compared to the algorithm that do not consider compatibility.

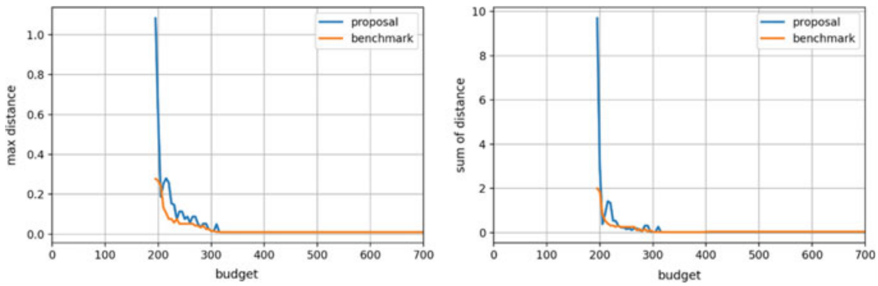
4.2 Experiment 2: Comparison with Algorithm from the Related Study

We performed experiments using $f_d(\mathcal{G})$ and $f_s(\mathcal{G})$ as evaluation indices with Rahman et al.’s algorithm for minimizing *DiaDist,* which is a measure of compatibility defined in Eq. (1). The $f_d(\mathcal{G})$ of the teams formed by the proposed algorithm is expected to be greater than of Rahman et al.’s algorithm based on full search. Since Rahman et al.’s algorithm cannot form teams for multiple tasks, in this experiment, teams are formed for only the first task used in Experiment 1. We randomly select 5,000 workers out of the 69,091 workers prepared for the experiment because the computational complexity is considerably large. In addition, the full search in the binary tree is conducted by depth-first search. We conducted an experiment using a fixed budget and skill level as a variable parameter, and fixed skill level and budget as a variable parameter.



(1) maximum distance (2) sum of distance

Fig. 1 Comparison to Rahman et al.'s algorithm with the infinite budget for each skill level (Ex. 2)



(1) maximum distance (2) sum of distance

Fig. 2 Comparison to Rahman et al.'s algorithm with skill level fixed to 15 for each budget (Ex. 2)

The results of the comparison experiment with infinite budget and skill levels in the range of $\{1, 2, \ldots, 100\}$ are shown in Fig. 1, and the results of the comparison experiment with a fixed skill level of 15 and a budget in the range of $\{50, 55, \ldots, 700\}$ are shown in Fig. 2.

According to Fig. 1(1), we can observe that the proposed algorithm becomes unstable for the small range of skill levels, but remains unchanged for most points. The maximum difference is approximately 0.04, and it is significantly small considering that the median of the collaboration costs by all worker combinations is approximately 5. Moreover, there are a few areas where the max distance on team of teams formed by proposed algorithm is smaller than one by Rahman et al's. In both algorithms, the subgroups are formed from workers whose distance from the root worker is less than the distance threshold τ. Rahman et al.'s algorithm, which performs all search, finds the optimal solution by minimizing τ. The actual maximum distance on team varies between τ and 2τ. Hence, this reversal phenomenon is considered to have occurred.

The comparison of the sum of distances on team indicates that the proposed algorithm is improved compared with Rahman et al.'s when the skill levels are high as a result in Fig. 1(2). The number of members in a subgroup becomes large under the condition of high skill level, and the total distance in the team increases. The proposed algorithm reduces the number of team members by exchanging workers to ensure that the sum of distances on team is smaller.

Conversely, the maximum distance on team and the sum of distances on team of teams formed by Rahman et al.'s algorithm are smaller, especially for a range of small budget as a result in Fig. 2. In the experiment with the budget as the parameter, there are a few worker exchanges processes because the experiment is performed under low skill level conditions and limited budget. As a result, the value of Rahman et al.'s algorithm is smaller for most of the range of conditions. However, these differences are not expected to have a significant effect similar to that shown in Fig. 1.

5 Conclusion

In this study, we proposed a system to construct an organization that executes complex tasks by considering compatibility among crowd workers. We have confirmed that the proposed team formation algorithm significantly reduces the maximum distance and the sum of distances on team for teams with the smallest budget for each skill requirement compared with that of the algorithms that do not consider compatibility. While forming a single team, the maximum distance on team and the sum of distances on team are larger than that in the Rahman et al.'s algorithm under strict budget constraints. However, the differences might not have a significant impact compared with that of the median of collaboration costs. Moreover, the sum of distances is smaller when the budget is available. In our future work, we plan to conduct research focusing on the processing of new workers that do not exist in the SN and time series change due to team formation.

Acknowledgements This work was supported by JSPS KAKENHI Grant Numbers JP18K18159, JP21H03553.

References

1. Kashima, H., Oyama, S., Baba, Y.: Human Computation and Crowdsourcing. Kodansha (2016)
2. Hertel, G.: Synergetic effects in working teams. J. Manag. Psychol. **26**(3), 176–184 (2011). https://doi.org/10.1108/02683941111112622
3. Huffmeier, J., Hertel, G.: When the whole is more than the sum of its parts: group motivation gains in the wild. J. Exp. Soc. Psychol. **47**(2), 455–459 (2011). https://doi.org/10.1016/j.jesp.2010.12.004
4. Chen, S.-J., Lin, L.: Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. IEEE Trans. Eng. Manag. **51**(2), 111–124 (2004). https://doi.org/10.1109/TEM.2004.826011

5. Valentine, M.A., Retelny, D., To, A., Rahmati, N., Doshi, T., Bernstein, M.S.: Expert crowdsourcing with flash teams. In: 2017 CHI Conference on Human Factors in Computing Systems, pp. 3523–3537 (2017). https://doi.org/10.1145/2642918.2647409
6. Lappas, T., Liu, K, Terzi, E.: Finding a team of experts in social networks. In: 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 467–476 (2009). https://doi.org/10.1145/1557019.1557074
7. Gajewar, A., Sarma, A. D.: Multi-skill collaborative teams based on densest subgraphs. In: 2012 SIAM International Conference on Data Mining, pp. 165–176 (2012). https://doi.org/10.1137/1.9781611972825.15
8. Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., Leonardi, S.: Online team formation in social networks. In: 21st International Conference on World Wide Web, pp. 839–848 (2012). https://doi.org/10.1145/2187836.2187950
9. Wang, W., He, Z., Shi, P., Wu, W., Jiang, Y., An, B., Hao, Z.: Chen, B, Strategic social team crowdsourcing: forming a team of truthful workers for crowdsourcing in social networks. IEEE Trans. Mob. Comput. **18**(6), 1419–1432 (2019). https://doi.org/10.1109/TMC.2018.2860978
10. Rahmam, H., Roy, S. B., Thirumuruganathan, S., AmerYahia, S., Das, G.: Task assignment optimization in collaborative crowdsourcing. In: 2015 IEEE International Conference on Data Mining, pp. 949–954 (2015). https://doi.org/10.1109/ICDM.2015.119
11. Stack Overflow Developer Survey 2020. https://insights.stackoverflow.com/survey/2020#top-paying-technologies.
12. Employment-Hours worked-OECD Data. https://data.oecd.org/emp/hours-worked.htm

Matrix Approach and Scheduling for Cooperation Requirement Planning in Industrial Robots

Tomohiro Nakada

Abstract Recently, companies have been shifting from lean manufacturing to cell manufacturing using industrial robots. Nof et al. have proposed cooperation requirement planning (CRP) among multiple industrial robots and tasks in the production process. CRP comprises two parts, CRP-I and CRP-II. CRP-I establishes the conditions for ensuring cooperation among multiple industrial robots, whereas CRP-II formulates an operation plan for optimizing the manufacturing objectives using the conditions established by CRP-I. Nof's study describes a method for formulating a consistent operation plan based on a two-step method procedure where conditions are set by CRP-I and optimized by CPR-II in multiple industrial robots. Furthermore, the matrix approach is a method of representing the relationship between inputs and outputs and examining operational plans. The present study focuses on CRP with multiple industrial robots and proposes a mathematical model using a matrix approach (Matsui's equation) that includes scheduling method. The proposed mathematical model demonstrates the possibility of reducing the abovementioned two operation steps (CRP-I and CRP-II) to one and includes the scheduling theory by adding one additional matrix to Matsui's equation. Furthermore, the mathematical model can compare the corrected values obtained for two cases involving CRP with multiple industrial robots based on this modified matrix approach with the standard values obtained from the original approach. The matrix approach examines the operation plans of multiple industrial robots by comparing the standard values and corrected values. This paper presents a mathematical model and development from the conventional two-stage method (CRP-I and CRP-II) to a one-stage method (matrix approach) of Matsui based on the cooperation requirement planning of multiple industrial robots.

Keywords Matrix approach · Cooperation requirement planning · Scheduling

T. Nakada (✉)
Bunkyo Gakuin University, Tokyo, Japan
e-mail: tnakada@bgu.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_3

1 Introduction

Companies have been recently shifting from lean manufacturing to cell manufacturing because of consumer diversity and short usage periods. Industrial robots are among the tools used in cell manufacturing, including both conventional industrial robots and high-performance robots equipped with the Internet of Things (IoT) technology [1]. Cell manufacturing involves combined work between humans and industrial robots and the development of efficient plans for the utilization of multiple industrial robots [2, 3]. This study focuses on task planning and scheduling for multiple industrial robots in such situations. Nof et al. proposed cooperation requirement planning (CRP), which considers cooperation among multiple industrial robots and the tasks involved in the production process [4–7].

Conversely, Matsui proposed a matrix approach (Matsui's equation) [8] as a model for realizing the goals of a company. It uses matrix products to express the relationships between inputs and outputs and their conversions within a company. The matrix approach presents elements related to management resources, such as human resources, goods, capital, and information, in rows and columns. Furthermore, Matsui has proposed a theory of modern economic growth toward facilitating a shared society (or creating shared value) that employs the scheduling theory with a focus on prioritizing rules [9]. Scheduling theory involves describing a process of decision-making used by many manufacturing and service industries to optimize one or more objectives [10, 11]. The typical goal of optimization is to minimize the time taken to complete the last task in the production plan by prioritizing rules. Studies using the rule of the longest processing time (LPT) have developed integer linear-programming models and heuristic algorithms to minimize the timespan required to make a product (make-span) using identical parallel machines [12]. Therefore, a matrix approach (Matsui's equation) that uses the LPT rule to minimize the make-span is examined herein.

A mathematical model that combines the matrix approach and scheduling for CRP among industrial robots to operate multiple industrial robots for completing specific tasks within a given job is proposed herein. The rest of this paper is organized as follows. Section 2 discusses the CRP among industrial robots and also describes a scheduling theory as related research. Section 3 describes a mathematical model based on a previous study that uses CRP for industrial robots. Section 4 compares and discusses the results of calculations that assume changes in the distribution of tasks and industrial robots. Finally, we conclude this study in Sect. 5.

2 Related Work

2.1 Cooperation Requirement Planning Among Industrial Robots

The conceptual basis of previous studies involves a cell production system that uses industrial robots [4, 5]. Herein, a planning environment in which multiple industrial robots are placed around a central worktable is considered. A single task can use either one industrial robot or multiple industrial robots. A "resource" is defined either as a single industrial robot or a combination of multiple industrial robots. Previous research proposed a planning method based on industrial robots, resources, and tasks as CRP.

Furthermore, it is necessary to review the task plan to accommodate the maintenance and breakdown of industrial robots. Multiple industrial robots can replace other conventional industrial robots. Accordingly, this paper examines a planning method based on industrial-robot resources and tasks in a cell production process that employs multiple industrial robots.

Nof et al. have proposed CRP based on multiple industrial-robot tasks and available resources. This CRP comprises two parts (CRP- I and CRP- II), as shown in Fig. 1. CRP-I defines single and multiple combinations that use multiple industrial robots as resources. CRP-II selects a plan to satisfy each goal and resolve the interactions between multiple industrial robots.

2.2 Matrix Approach

The matrix approach is a system-design method that represents the systematization of human resources, goods, cost, and information to support decision-making. Matsui's matrix equation represents the input and output relationships for decision-making within a company [8]. The matrix approaches can be classified into two types, as shown in Fig. 1: the structural-matrix method (tabular form) and Matsui equation (compact form). The input and output relationships of the company add two matrices—Balancing (B) and Goal (G)—to create a four-part logical structure (Ki-sho-ten-ketsu is a type of structure in Japanese writing) as follows: I, Introduction (Ki); D, Development (sho); T, Transformation (ten); and C, Conclusion (ketsu).

Matsui's matrix equation provides an approach that supports decision-making. Therefore, mathematical models based on Matsui's equation are examined herein to implement CRP for industrial robots, as proposed by Nof et al. [4, 5].

$$\text{Objective form: } \mathrm{g} = \mathrm{I}^{\mathrm{T}}\mathrm{D}^{\mathrm{T}}\mathrm{T}^{\mathrm{T}}\mathrm{C}^{\mathrm{T}}\mathrm{B}^{\mathrm{T}} \quad (1)$$

$$\text{Unique form: } \mathrm{I}^{\mathrm{T}}\mathrm{D}^{\mathrm{T}}\mathrm{T}^{\mathrm{T}}\mathrm{C}^{\mathrm{T}}\mathrm{g} = \lambda g \quad (2)$$

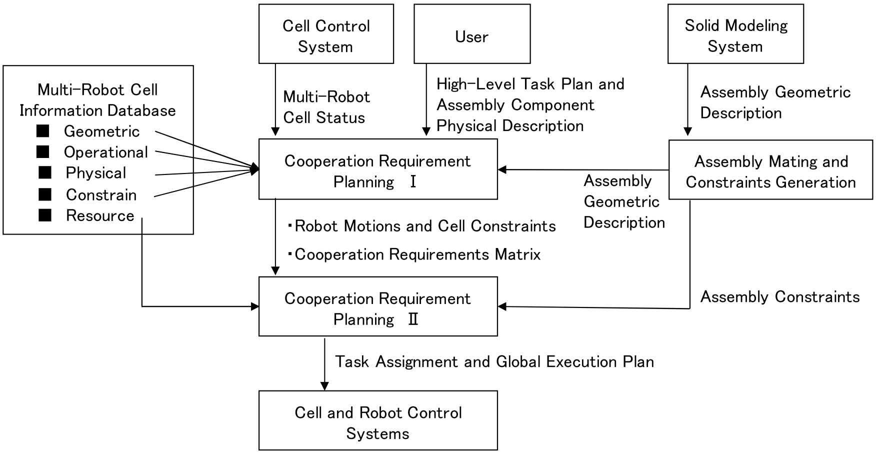


Fig. 1 Cooperation-requirement-planning system architecture [5]

Figure 1 represents Matsui's matrix equation in the format I → D → T → C → B → G. The inverse matrix has the form G → B → C → T → D → I. Matsui's equation can be represented either in the following objective form (Eq. 1) or in the unique form provided in Eq. 2.

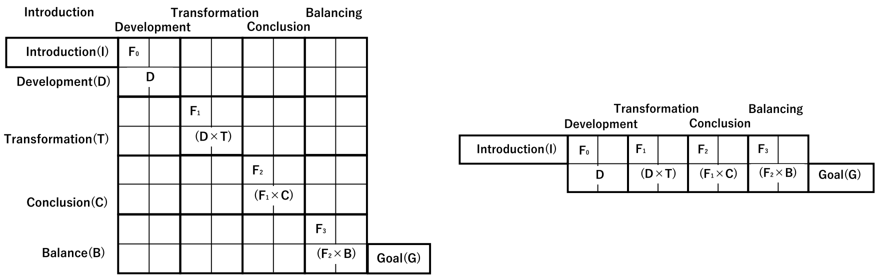
2.3 Scheduling Theory

Matsui [9] proposed the longest processing time (LPT) rule based on the principle of balancing vs. sharing trade-offs in an autonomous economy under manufacturing logic. Here, the mean flow time and maximum flow time is indicated based on job-shop scheduling of the sequencing type [11]. Scheduling problems with *m* machines and *n* jobs are often abbreviated as type n/m/Fmax, where Fmax is the maximum flow time. For example, two machines and *n* jobs describe the Johnson rule, which is abbreviated as n/2/Fmax [11].

3 Mathematical Model

A mathematical model based on Matsui's matrix equation for CRP among industrial robots to plan the coordinated operation of multiple industrial robots is proposed herein, as shown in Fig. 2. The equation shown in this figure expresses the resources and scheduling of multiple industrial robots and yields the maximum residence time (make-span) and the mean flow time for a particular job.

The constraints on this equation are as follows:



(a) Table form of the equation

(b) Compact form of the equation

Fig. 2 Two types forms: **a** structural table and **b** Matsui's compact matrix equation

- Multiple industrial robots are used.
- Each job is processed according to the processing procedure.
- Each resource is either a single task performed by an industrial robot or is a cooperative requirement task that uses multiple industrial robots.
- The time for each processing step is the same, and each step has the same start and end times.
- There is no idle time in any processing step.
- Industrial robots process all work without any interruption.

This section focuses on CRP [4, 5] and describes the detailed procedure leading up to the construction of the equation shown in Fig. 2. The operation plans for CRP-I and CRP-II are expressed using Matsui's equation, as shown in Fig. 3. The left side of Fig. 3 illustrates CRP-I, where ROB indicates an industrial robot and "r" indicates a resource. For example, "r1" represents a single operation because only ROB1 is defined. Conversely, "r4" is defined as a combination of ROB1 and ROB2 and is a cooperative request. The right side of Fig. 3 illustrates CRP-II, which represents the relationship between tasks ("t1" to "t3") and resources ("r1" to "r7") using a cooperative requirement matrix. Herein, "t" represents the processing time rather than a given task. This matrix represents the use of a resource with the entry 1 and no use of that resource with 0. Thus, Nof's cooperative request approach is represented as a planning method comprising the two stages CRP-I and CRP-II (Fig. 4).

The Introduction in Fig. 2, which is based on Matsui's equation, represents industrial robots. The requirement for cooperation among multiple industrial robots is determined by the product of Introduction (I) and Development (D). The relationship between resources and tasks is determined by the product of Development (D) and Transformation (T). This matrix approach is illustrated by the examples discussed below.

In addition, Conclusion (C) is a matrix that transforms resources into time units to introduce the scheduling theory. Finally, balancing (B) is a matrix that represents the mean flow time and priority scheduling (the LPT rule) for each process illustrated in Fig. 2 The equation represented in Fig. 2 can be used to calculate the make-span (F_I) and mean flow time (F_{II}) for multiple industrial robots and tasks.

Introduction

$$\begin{pmatrix} \overset{ROB1}{1} & \overset{ROB2}{1} & \overset{ROB3}{1} \end{pmatrix} \times$$

Development Transformation Conclusion Balancing

$$\begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix}\begin{pmatrix} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \\ 1 & & & 1 & 1 & & 1 \\ & 1 & & 1 & & 1 & 1 \\ & & 1 & & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} t_1 & t_2 & t_3 \\ 1 & & \\ 1 & & \\ 1 & & \\ 1 & 1 & \\ 1 & 1 & \\ 1 & 1 & \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \bar{L} & \bar{Z} \\ 1/7 & 7 \\ 1/7 & 6 \\ 1/7 & 5 \\ 1/7 & 4 \\ 1/7 & 3 \\ 1/7 & 2 \\ 1/7 & 1 \end{pmatrix}$$

Goal

$$= (F_{II} \quad F_{I})$$

Fig. 3 CRP using the matrix approach and scheduling

Fig. 4 Resource–robot relationship diagram (CRP-I, left) and resource–task relationship diagram (CRP-II, right)

r1={ROB1}
r2={ROB2}
r3={ROB3}
r4={ROB1, ROB2}
r5={ROB1, ROB3}
r6={ROB2, ROB3}
r7={ROB1, ROB2, ROB3}

	t1	t2	t3
r1	1	0	0
r2	1	0	0
r3	1	0	0
r4	1	1	0
r5	1	1	0
r6	1	1	0
r7	1	1	1

4 Comparison of Calculation Results

This section performs calculations using the proposed mathematical model and compares various numerical results. First, Eq. 3 is used to calculate the numerical value for an application example of CRP. The mean flow time and make-span result obtained by matrix multiplication are 24 and 672, respectively, which are as “standard values.” Furthermore, the production plans for two cases resulting for CRP from for industrial robots based on these two values are examined herein.

Case 1 assumes that goods are manufactured by changing the order of resources and tasks. In other words, it is a modification of the Development and Transformation components of the mathematical model. Equation 4 below shows the results calculated with the beginnings and ends of the task changed. The combination of resources is changed based on the order of the tasks. The mean flow time and make-span result obtained by matrix multiplication are 24 and 672, respectively. As the results of calculations using Eqs. 3 and 4 both yield the same value, changing the tasks and resources does not cause any outcome.

Case 2 assumes that ROB2 is broken. In other words, it is a modification of the Introduction factor of the mathematical model (the second entry from the left in that factor). Equation 5 shows the results of a calculation that assumes ROB2 = 0. The mean flow time and make-span obtained by matrix multiplication are now 16 and 448, respectively, which are lower than the standard values.

$$\begin{bmatrix}1&1&1\end{bmatrix}\begin{bmatrix}1&0&0&1&1&0&1\\0&1&0&1&0&1&1\\0&0&1&0&1&1&1\end{bmatrix}\begin{bmatrix}1&0&0\\1&0&0\\1&0&0\\1&1&0\\1&1&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\end{bmatrix}\begin{bmatrix}\frac{1}{7}&7\\\frac{1}{7}&6\\\frac{1}{7}&5\\\frac{1}{7}&4\\\frac{1}{7}&3\\\frac{1}{7}&2\\\frac{1}{7}&1\end{bmatrix}=\begin{pmatrix}24&672\end{pmatrix} \quad (3)$$

$$\begin{bmatrix}1&1&1\end{bmatrix}\begin{bmatrix}1&0&0&1&1&0&1\\1&1&0&1&0&1&0\\1&0&1&0&1&1&0\end{bmatrix}\begin{bmatrix}1&1&1\\1&0&0\\1&0&0\\1&1&0\\1&1&0\\1&1&0\\1&0&0\end{bmatrix}\begin{bmatrix}1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\end{bmatrix}\begin{bmatrix}\frac{1}{7}&7\\\frac{1}{7}&6\\\frac{1}{7}&5\\\frac{1}{7}&4\\\frac{1}{7}&3\\\frac{1}{7}&2\\\frac{1}{7}&1\end{bmatrix}=\begin{pmatrix}24&672\end{pmatrix} \quad (4)$$

$$\begin{bmatrix}1&0&1\end{bmatrix}\begin{bmatrix}1&0&0&1&1&0&1\\0&1&0&1&0&1&1\\0&0&1&0&1&1&1\end{bmatrix}\begin{bmatrix}1&0&0\\1&0&0\\1&0&0\\1&1&0\\1&1&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\end{bmatrix}\begin{bmatrix}\frac{1}{7}&7\\\frac{1}{7}&6\\\frac{1}{7}&5\\\frac{1}{7}&4\\\frac{1}{7}&3\\\frac{1}{7}&2\\\frac{1}{7}&1\end{bmatrix}=\begin{pmatrix}16&448\end{pmatrix} \quad (5)$$

5 Conclusion

This study focuses on CRP with multiple industrial robots and proposes a mathematical model using a matrix approach (Matsui's equation) that includes scheduling. The proposed mathematical model demonstrates the possibility of reducing two operation plans (CRP-I and CRP-II) to one operation plan, and it also incorporates the scheduling theory by introducing balancing into matrix approach used for this case study.

This study also illustrates the results of calculations ("standard values") obtained using the CRP proposed by Nof et al. In addition, the results are calculated for two illustrative cases (change of task order, failure of one industrial robot), and these are compared with the standard values. If the task-order change is paired with the resource change, the mean flow time and make-span are the same as the standard values. Conversely, the failure of one industrial robot reduces both resources and tasks, resulting in the mean flow time and make-span being less than the standard values. Therefore, the proposed mathematical model—based on a matrix approach that considers CRP with multiple industrial robots—can be used to obtain the corrected values by comparison with the standard values.

References

1. Shariatzadeh, N., Lundholm, T., Lindberg, L., Sivard, G.: Integration of digital factory with smart factory based on Internet of Things. Procedia CIRP **50**, 512–517 (2016)
2. Matheson, E., Minto, R., Zampieri, E.G.G., Faccio, M., Rosati, G.: Human-robot collaboration in manufacturing applications: a review. Robotics **8**(4), 100 (2019). https://doi.org/10.3390/robotics8040100
3. Perez, L., Rodrigues-Jimenez, S., Rodrigez, N., Usamentiaga, R., Garcia, D.F.: Digital twin and virtual reality based methodology for multi-robot manufacturing cell commissioning. Appl. Sci. **10**(10), 3633 (2020). https://doi.org/10.3390/app10103633
4. Nof, S.Y., Rajan, V.N., Lenz, E.: Automatic generation of assembly constraints and cooperation task planning. CIRP Ann. **42**(1), 13–16 (1993)
5. Rajan, V.N., Nof, S.Y.: Cooperation requirements planning (CRP) for multiprocessors: optimal assignment and execution planning. J. Intell. Rob. Syst. **15**, 419–435 (1996)
6. Nof, S.Y., Ceroni, J., Jeong, W., Moghaddam, M.: Revolutionizing Collaboration through e-Work, e-Business, and e-Service, Automation, Collaboration, and E-service Book Series. Springer, Berlin (2015). https://doi.org/10.1007/978-3-662-45777-1
7. Zhong, H., Nof, S.Y., Berman, S.: Asynchronous cooperation requirement planning with reconfigurable end-effectors. Rob. Comput. Integr. Manuf. **34**, 95–104 (2015)
8. Matsui, M.: Fundamentals and Principles of Artifacts Science: 3M&I-Body System, Springer, Berlin (2016)
9. Matsui, M.: A theory of modern economic growth toward sharing society. Theor. Econ. Lett. **8**(4), 675–682 (2018)
10. Pinedo, M.L.: Scheduling theory, algorithms, and systems, 4th edn. Springer, Berlin (2012)
11. Hitomi, K.: Manufacturing systems engineering: a unified approach to manufacturing technology, production management and industrial economics, 2nd edn. Routledge, (2017)
12. German, Y., Badi, I., Bakir, A., Shetwan, A.: Scheduling to minimize makespan on identical parallel machines. Int. J. Sci. Eng. Res. **7**(3), 353–359 (2016)

Evaluation of Material-Based GHG Emissions Under COVID-19 Disruption on Redesigning Global Supply Chain Network Across TPP Countries

Takaki Nagao, Hiromasa Ijuin, Keisuke Nagasawa, and Tetsuo Yamada

Abstract In recent years, global warming has become a serious problem in a global supply chain which is a series of cross-border transaction including custom duty that is a tax imposed for imported goods. In order to prevent the global warming, Green House Gas (GHG) emissions needs to be reduced throughout the supply chain. Moreover, procurement costs, material-based GHG emissions, tariff in countries are different by each country. In addition, the Trans-Pacific Partnership (TPP), which is a free trade agreement signed between 11 countries, including Japan and Malaysia, have promoted the trade in parts and products among the TPP participant countries without customs duty. Thus, the network on the global supply chain affects not only costs but also GHG emissions. On the other hand, the disruption by COVID-19 caused adverse impacts to redesign supply chains all over the world, where parts or materials are not provided from current suppliers by the disruption. Thus, the network may be reconfigured, which brings different GHG emissions in the global supply chain network before and after the disruption across TPP countries. The purpose of this study is to evaluate material-based GHG emissions under COVID-19 disruption on redesigning global supply chain network across TPP countries. Firstly, global supply chain network is modeled and formulated. Next, numerical experiments are conducted for evaluating material-based GHG emissions under disruption scenarios. Finally, the results are analyzed in terms of GHG emissions and costs. The result shows that the highest reduction ratio of the total GHG emission on a global supply chain is 58.4% compared to the baseline.

Keywords Green supply chain · Economic partnership · Lifecycle inventory database · Bill of Materials · 0–1 integer programming

T. Nagao · H. Ijuin · T. Yamada (✉)
The University of Electro-Communications, Tokyo, Japan
e-mail: tyamada@uec.ac.jp

T. Nagao
e-mail: n2130089@edu.cc.uec.ac.jp

K. Nagasawa
Hiroshima University, Hiroshima, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_4

1 Introduction

The scale of disruption by COVID-19 has been impacted on supply chains all over the world. According to Min et al. [1], 53% of active food suppliers in a month did not have enough food sources as compared to previous years by the preventive measures for disruption of COVID-19. Additionally, 83.1% of food suppliers' revenue was decreased. A global supply chain network is a series of procurement, production, storage, transportation, and sales across national borders [2]. In the cross-border transactions, imports are imposed a custom duty, which is a kind of border tax adjustment for importing [3]. The Trans-Pacific Partnership (TPP) is a free trade agreement signed between 11 countries, including Japan and Malaysia. The TPP made the trade in parts and products between the TPP participant countries free of customs duty since 2020 [4]. Additionally, disruption is defined as a situation where the base disconnected to a supply network in a consequence of a disruption [5].

On the other hand, though lockdowns by COVID-19 stagnated economy and reduced CO_2 emission by 7% in 2020, the emission has increased once again [6]. Global warming is a serious problem that greenhouse gases (GHG) cause increment of temperature on the earth. A major climate summit was held on April 22 in 2021, and US President Joe Biden promised to reduce carbon emissions by 50–52% comparing to 2005 levels by 2030 [7]. In order to build a global low-carbon supply chain in the manufacturing sector, estimating and visualizing the reduction of GHG emissions need the entire supply chain including the Asian countries [8] because Asian countries, China, and the United States occupied GHG emission of the world by 41.9%, 28.4%, 14.7%, respectively [9]. Therefore, it is necessary to evaluate GHG emission, COVID-19, and TPP on a global supply chain simultaneously. On a supply chain, GHG emission by manufacturing components is 95.5% [10]. It is noted that countries have different material-based GHG emission and costs even when manufacturing the same component. Kondo et al. [11] conducted numerical experiments, and the results indicated that different carbon tax induced carbon leakage on the global supply chain.

In previous studies, a global review and supply chains analysis in macro level on a field of construction industry focusing on carbon footprint were conducted by Nuri et al. [12] using Scopus database from 2009 to 2020. They pointed out that Life Cycle Assessment (LCA) had been used in only 10% of the literatures. Furthermore, Hayashi et al. [13] designed a global supply chain network across TPP countries with carbon tax based on the material-based emissions for components manufacturing. However, these studies did not consider GHG emission, disruptions and TPP simultaneously.

The purpose of this study is to evaluate the changes of the material-based GHG emissions in a reconfiguring global supply chain with and without TPP under the COVID-19 disruptions. The rest of this paper is structured as follows. In Sect. 2, a model and formulations of a global supply chain with TPP under disruptions are described. Additionally, it is explained how the material-based GHG emissions inside parts procured are calculated using LCA. After that, the assumptions and disruption scenarios for numerical experiments are given in Sect. 3. Section 4 shows and discusses the impact on the GHG emissions and costs across the TPP and non-TPP countries. Finally, Sect. 5 concludes this study and mentions future studies.

2 Model and Formulation

2.1 Model

The TPP and a global supply chain model with supply disruption used in this study is based on Nagao et al. [14]. Figure 1 shows a global supply chain network model under supply disruptions with TPP to evaluate the material-based GHG emissions and costs. The material-based GHG emission for each part in each country manufactured in this study are estimated with material procurement cost by GHG emission level in the country listed in the life cycle inventory database with the Asian International Input and Output Tables [8] as well as Kondo et al. [11]. Here, part j is produced by supplier o and transported to factory p. Then, the finished product, consisting of N_j parts, is assembled in factory p. Finally, finished products are transported to the market q.

The global supply chain network with supply disruption and TPP is modeled in this study as follows. Given that the supplier cities face the possibility of the disruption, it is assumed that supply parts are impossible to a factory if the supplier city suffers the disruption. In this model, the suppliers from disrupted supplier cities

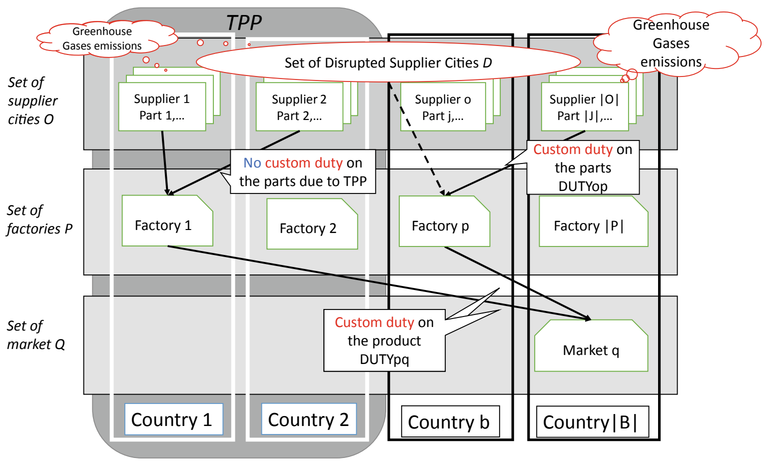


Fig. 1 Global supply chain network model with TPP under the supply disruptions and material-based GHG emissions

represented as set D cannot supply all parts to any factory. Therefore, the parts should be transported from non-disrupted suppliers.

Customs duties are imposed when parts or products are imported across the country or TPP borders. The material-based GHG emissions for parts are occurred by manufacturing parts at supplier o, thus, the emission is according to the country which supplier o is located in. The reason why GHG emission of parts manufacturing is focused on in this study is that GHG emitted in a process of manufacturing parts by 95.5% [10]. With respect to TPP and custom duties, suppliers, factories, and markets are located in a set of countries B. In this model, group G defines a set of supplier cities in TPP countries.

The notations used in this study are as follows:

Sets:

B Set of countries, $b \in B$
O Set of suppliers, $o \in \mathrm{O}$
G Set of suppliers on countries forming the Trans-Pacific Partnership Agreement, $g \in \mathrm{G}$
D Set of disrupted supplier cities in supplier set O, $d \in D$, $D \subset O$
J Set of parts, $j \in J$
P Set of factories, $p \in P$
Q Set of markets, $q \in \mathrm{Q}$

Decision variables:

v_{opj} Amount of parts j transported from supplier o to factory p
l_{oj} Amount of part j produced at supplier o
v_{pq} Amount of products transported from factory p to market q
k_p Amount of products manufactured in factory p
z_{pq} 1 if the route between factory p and market q is opened, and 0 otherwise
u_p 1 if factory p is opened, and 0 otherwise

Cost parameter:

C_{op}^{LC} Logistics cost per unit part of transporting from supplier o to factory p
C_{pq}^{LC} Logistics cost per unit product of transporting from factory p to market q
C_{oj}^{PC} Procurement cost of procuring per unit part j by supplier o
C_{op}^{T} Customs duty per unit on transportation from supplier o to factory p
C_{pq}^{T} Customs duty per unit on transportation from factory p to market q
C_{p}^{M} Manufacturing cost per product at factory p
C_{pq}^{R} Fixed cost for opening route between factory p and market q
C_{p}^{F} Fixed cost for opening factory p

Production parameter:

N_j Total number of part j consisting of one product
$N_{product,q}$ Amount of demand for product at market q
M Very large number (Big M)

C_p Production capacity at factory *p*
S_{oj} 1, if part *j* is supplied by supplier *o*, and 0 otherwise

Cost evaluation:

TC Total cost [USD]
TMC Total manufacturing cost [USD]
TTC Total transportation cost [USD]
TCDC Total custom duty cost [USD]

GHG Emission evaluation:

H_{oj} GHG emission of part *j* for manufacturing country *b* at supplier *o* based on Life Cycle Assessment (LCA) [g-CO_2eq]
TG Total material-based GHG emission [g-CO_2eq]

2.2 Formulation

The global supply chain network is formulated as an Integer Programming (IP) [15] based on Hayashi et al. [13] and Nagao et al. [14]. It is known that IP can treat and solve transportation and assignment problems as linear programming [15]. The objective function of this study is formulated to minimize total cost *TC,* consisted of total manufacturing cost *TMC*, total transportation cost *TTC* and total custom duty cost *TCDC* as shown in Eq. (1). *TMC* is the sum of procurement cost of parts, manufacturing cost of products, route opening cost and factory opening cost. *TTC* is the sum of transportation cost of parts from suppliers to factories and products from factories to markets. *TCDC* is the sum of custom duty cost of parts and products.

Objective function:

$$TC = TMC + TTC + TCDC \rightarrow \min \tag{1}$$

$$TMC = \sum_{o\in O}\sum_{p\in P}\sum_{j\in J} C_{oj}^{PC} v_{opj} + \sum_{p\in P}\sum_{q\in Q} C_p^M v_{pq} + + \sum_{p\in P}\sum_{q\in Q} C_{pq}^R z_{pq} + \sum_{p\in P} C_p^F u_p \tag{2}$$

$$TTC = \sum_{o\in O}\sum_{p\in P}\sum_{j\in J} C_{op}^{LC} v_{opj} + \sum_{p\in P}\sum_{q\in Q} C_{pq}^{LC} v_{pq} \tag{3}$$

$$TCDC = \sum_{o\in O}\sum_{p\in P}\sum_{j\in J} C_{oj}^{PC} C_{op}^T v_{opj} + \sum_{p\in P}\sum_{q\in Q} C_p^M C_{pq}^T v_{pq} \tag{4}$$

Equations (5) and (6) are defined to estimate the material-based GHG emissions. The other constraints are set similar to Nagao et al. [14]. *TG* is the sum of GHG emissions by multiplying parts and GHG emission by each country.

$$\sum_{p \in P} v_{opj} = l_{oj} \quad \forall o \in O, \quad \forall j \in J \tag{5}$$

$$TG = \sum_{o \in O} \sum_{j \in J} H_{oj} l_{oj} \tag{6}$$

3 Assumptions and Scenarios

3.1 Assumptions

To analyze the changes of GHG emissions caused by COVID-19 supply disruptions in countries with and without TPP in a global supply chain, this study considers a vacuum cleaner with 23 types of parts as an example product, as well as Nagao et al. [14]. However, it is assumed that a part of motor is excluded because the procurement cost of the motor is about 95% of the total cost procured for all the parts. The other assumptions in a global supply chain is prepared in this study as follows.

- Fifty-two cities, that is, 13 each from China, Malaysia, the United States, and Japan, are selected as suppliers.
- Four cities, Shanghai, Kuala Lumpur, Seattle and Tokyo, are selected as candidates for a factory location. The production capacity of each factory is 3,000 units, and the market demand set at Seattle, the United States, is 6,000 units.
- The customs duty rate between the United States and China is 25%, and that between a TPP country and a non-TPP one is 10%. For instance, the customs duty rate between Malaysia and China and between Japan and the United States becomes 10%. Meanwhile, there is no customs duty between TPP countries, such as Malaysia and Japan.
- In a supplier disruption, the supply of parts from the supplier to factories is disrupted. Therefore, the parts are transported from non-disrupted suppliers.

The parts of a product have different emissions and costs in different countries. For example, the GHG emissions of a part #16 upper filter are 536.47 [g-CO_2eq] in the U.S., 1,164.65 [g-CO_2eq] in Malaysia, 2,853.34 [g-CO_2eq] in China, and 421.36 [g-CO_2eq] in Japan. Procurement costs of it are 0.3085 [USD] in the U.S., 0.2535 [USD] in Malaysia, 0.2782 [USD] in China, and 0.5759 [USD] in Japan.

3.2 Disruption Scenarios

In this study, two disruption scenarios are considered similar to Nagao et al. [14]. Disruptions are assumed as a decision of each government whether a city is disrupted or not. Additionally, it is assumed that the disruption area by COVID-19 is increasing from scenario 1 to 3, and it would finally cover the whole country as follows. For example, scenario A with supplier disruption in China, a non-TPP country is developed.

Firstly, in scenario A1, the disruption is caused in three cities (Xian, Chengdu, and Chongqing). Next, in addition to scenario A1, the other 4 cities (Nanjin, Hangzhou, Jinan, and Suzhou) are also disrupted in scenario A2. Finally, in scenario A3, all cities in China are disrupted. Moreover, scenarios AB are also prepared to analyze the combination impact by scenarios A and B simultaneously.

- Baseline: No disruption.
- Scenario (A) Supplier disruption in China, a non-TPP country; A1, A2, A3: Disruption in 3, 7, and all cities in China.
- Scenario (B) Supplier disruption in Malaysia, a TPP country B1, B2, B3: Disruption in 3, 7, and all cities in Malaysia.
- Scenario (AB) Scenarios A and B combined. (A1B1, A2B2, A3B3).

4 Result

Figure 2 shows the result of the amount of GHG emissions and total cost in each scenario. Scenario B3 with all Malaysian suppliers disrupted results in the highest increment of GHG emission. The emissions are higher by 43.5% from the baseline, where the unit ratio of increment GHG emission for total cost is 3.6 [g-CO_2eq/USD]. In contrast, the largest reduction for GHG emission is ovserved in scenario A3B3 with all Malaysian and Chinese suppliers disruption, and the cost effectiveness for GHG reduction becomes 3.5 [g-CO_2eq/USD]. In scenario A3, the cost effectiveness for GHG reduction is 18.7 [g-CO_2eq/USD]. Therefore, scenario A3 is in superior to scenario A3B3 in terms of efficiency for GHG reduction and cost. As the degrees of disruption becomes larger in China, the amount of GHG emission is monotonously decreased by 4.7%, 35.9% and 58.4%, respectively. This is because the disruption of suppliers in China leads to the relocation of suppliers to Malaysia and Japan. This is because China has the highest GHG emission compared to Malaysia and Japan.

On the other hand, in scenario B, where Malaysian cities are disrupted, GHG emissions will increase due to the suppliers relocation to China. The reason for this increment of GHG emissions is that GHG emission for parts manufactured in Malaysia is the second highest next to the Chinese one. Additionally, the procurement cost for parts in China is the lowest among all countries. In scenario A2B2 and A3B3, the amount of GHG emission is also decreased by 56.6% and 73.7%, respectively. However, in scenario A1B1, the amount of GHG emission is increased by merely 4.1%. One of the reasons for this increment of GHG emission is that switching suppliers from Malaysia to China at Kuala Lumpur due to supplier disruptions in Malaysia exceeds the reduction of GHG emission, by switching suppliers from China to Japan at Shanghai caused by the disruptions in China. As the scale of the disruption

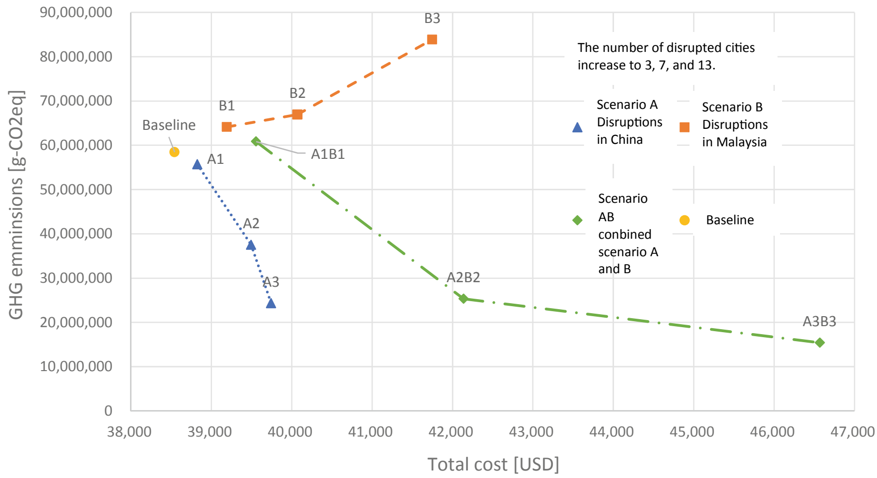


Fig. 2 The result of the amount of GHG emissions and total cost in each scenario

has been larger, the total cost has been increased. Additionally, the impact on the amount of GHG emissions becomes large.

The result indicated that when Chinese suppliers were disrupted, the total of GHG emission was decreased by switching suppliers from China to Malaysia and Japan, which brought less GHG emissions than Chinese one. The highest reduction ratio of the total GHG emission was 58.4% compared to the baseline. Meanwhile, when Malaysian suppliers were disrupted, Chinese suppliers were selected because of the low procurement cost and the highest increment of the total of GHG emission was 43.5% comparing to the baseline. Finally, when both Malaysian and Chinese suppliers were disrupted, GHG emission was greatly decreased when switching to Japanese and U.S. suppliers. Especially, when one of factory were relocated to U.S., suppliers were mainly switched to U.S. and the total of GHG emission was decreased by 56% or more. From the results, it is effective for reduction of GHG emission that a factory and suppliers are switched to a low GHG emission country near the market.

In scenario A3, the main reason for the increment of total cost is because the total transportation cost of products, and total customs duty of products are increased by USD 5,420, and 1,518, respectively. In scenario B3, the total transportation cost of products, the total procurement cost of parts, and the total customs duty of parts are increased by USD 2,125, 535, and 566, respectively. In scenario A3B3, the total transportation cost of products, total customs duty of products, and total procurement cost of parts are increased by USD 5,420, 5,910, and 2,764, respectively. Hence, it is found that the disruption of suppliers makes the breakdown of the total costs change.

5 Conclusion

Disruptions by COVID-19 influence the total of GHG emission on a global supply chain according to the GHG emission by procuring parts of disrupted country. In addition, 50% reduction of GHG emission is realized by relocating a factory from Shanghai to Seattle.

For future studies, redesigning a global supply chain network needs to include TPP and disruptions with bi-objective for GHG emission and costs, etc. Moreover, it should be taken into account to validate the proposed approach by comparing with actual behavior data. Additionally, the treated problem includes a political issue in the world stage because of international agreements that regulate GHG emission such as Paris agreement and Kyoto protocol. In that case, there may be adverse results, therefore, further study should guarantee that companies will be environmental and politics sensitive is to make it financially attractive as well, where business, environment and politics are intersected.

Acknowledgements This research was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Grant-in-Aid for Scientific Research (A), JP18H03824, from 2018 to 2023.

References

1. Min, S., Zhang, X., Li, G.: A snapshot of food supply chain in wuhan under the COVID-19 pandemic. China Agric. Econ. Rev. **12**(4), 689–704 (2020)
2. Nakamura, K., Ijuin, H., Yamada, T., Ishigaki, A., Inoue, M.: Design and analysis of global supply chain network with trans-pacific partnership under fluctuating material prices. Int. J. Smart Comput. Artif. Intell. **3**(1), 17–34 (2019)
3. Japan External Trade Organization (JETRO), Jetro Trade Handbook 2017, Japan External Trade Organization (JETRO) (2017). (In Japanese)
4. Ministry of Economy: Trade and Industry. Trans-Pacific Partnership (TPP). https://www.meti.go.jp/policy/external_economy/trade/tpp.html. Accessed on 31 May 2021
5. Kim, Y., Chen, Y.S., Linderman, K.: Supply network disruption and resilience: a network structural perspective. J. Oper. Manag. **33–34**, 43–59 (2015). https://doi.org/10.1016/j.jom.2014.10.006
6. CNN: Remember the pandemic greenhouse emissions drop? It's nearly gone. https://edition.cnn.com/2021/03/03/world/climate-greenhouse-gas-emissions-rebound-intl-scn/index.html. Accessed on 20 October 2022
7. BBC news: Biden: This will be 'decisive decade' for tackling climate change. https://www.bbc.com/news/science-environment-56837927. Accessed on 31 May 2021
8. Kokubu, K., Itsubo, N., Nakajima, M., Yamada, T.: Constructing low-carbon supply chain in Asia and the role of accounting. KAIKEI (ACCOUNTING) (2012). (In Japanese)
9. Ministry of the Environment: Global energy-related CO_2 emissions in 2018 (in Japanese). http://www.env.go.jp/earth/201222_co2_emission_2018.pdf. Accessed on 31 May 2021
10. SHARP: Environmental Report 2002. http://www.sharp.co.jp/corporate/eco/report/backnumber/pdf/esr2002e.pdf. Accessed on 31 May 2021

11. Kondo, R., Kinoshita, Y., Yamada, T.: Green procurement decisions with carbon leakage by global suppliers and order quantities under different carbon tax. Sustainability **11**(13) (2019). https://doi.org/10.3390/su11133710
12. Nuri, C.O., Murat, K.: Carbon footprint of construction industry: A global review and supply chain analysis. Renew. Sustain. Energy Rev. **124** (2020). https://doi.org/10.1016/j.rser.2020.109783
13. Hayashi, K., Matsumoto, R., Yamada,T., Nagasawa, K., Kinoshita, Y.: Designing a global supply chain network considering carbon tax and trans-pacific partnership agreement. In: Proceedings of the 2020 The society of Plant Engineering Japan Spring Conference, pp. 97–100 (2020). (In Japanese)
14. Nagao, T., Ijuin, H., Yamada, T., Nagasawa, K., Zhou, L.: COVID-19 Disruption strategy for redesigning global supply chain network across TPP countries. Logistics. 6(1), 2 (2021). https://doi.org/10.3390/logistics6010002
15. Hiller, F. S., Lieberman, G. J: Introduction to operations research. McGraw-Hill (2005)

Modeling of Inventory Routing Problem with Intermediate Locations in Emergency Logistics Considering Uncertainty of Road Conditions

Tomoki Sanada and Aya Ishigaki

Abstract In recent years, natural disasters have occurred worldwide, causing tremendous damage. It is difficult to predict the scale of damage caused by natural disasters before they occur; and even after the scale of damage is understood, it is difficult to make accurate recovery plans and decisions based on limited information. Therefore, it is necessary to construct a system that supports decision-makers. Emergency logistics is a new field of logistics that plays an important role in disaster relief and recovery. The inventory routing problem (IRP) in emergency logistics has been attracting attention to reduce damage caused by disasters. Unlike usual logistics, emergency logistics considers two uncertainties: the uncertainty of the demand scale and the uncertainty of the distribution network, such as roads. Oshima et al. (2020) proposed an IRP model that focuses on inventory management in secondary storage warehouses where processing operations are possible, in addition to demand points, assuming the delivery plans for materials such as lumber are planned in a situation where available distribution networks are known or have been restored. On the other hand, no model has been proposed that considers damage to distribution networks, such as roads, in the recovery phase. In this study, we propose an IRP model that considers the uncertainty of the road network and utilizes existing facilities as processing bases. In addition, numerical experiments were used to verify the effects of using processing bases in emergency logistics for machinable building components.

Keywords Inventory routing problem · Uncertainty of road conditions · Emergency logistics

T. Sanada · A. Ishigaki (✉)
Tokyo University of Science, Chiba, Japan
e-mail: ishigaki@rs.tus.ac.jp

T. Sanada
e-mail: 7421511@ed.tus.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_5

1 Introduction

Natural disasters occur worldwide and cause a great deal of damage. It is difficult to predict the scale of damage caused by natural disasters before they occur, and even after the scale of damage is known. Further, it is difficult to derive an accurate recovery plan and make decisions based on limited information at the disaster site. In recent years, the delivery planning problem called emergency logistics has been attracting attention with the aim of mitigating damage caused by disasters [1]. Unlike general business logistics, emergency logistics considers two uncertainties: the uncertainty of the demand scale and the uncertainty of the road condition [2, 3]. It is essential to consider the uncertainty of demand in emergency housing restoration because it takes more time to assess the extent of the damage, and accurate information on demand is not always available. In addition, it is difficult to maintain a balance between supply and demand because a delay in response can lead to a backlog of restoration requests. In other words, although an early response may shorten the lead-time for delivery and lead to a rapid restoration of housing, it is important to make decisions under conditions of high uncertainty; therefore, it is necessary to clearly present the evaluation and risks of each option to decision-makers.

Oshima et al. focused on the logistics of building materials such as lumber in the post-earthquake recovery phase and proposed a push-type delivery model based on predicted demand [4]. They used stochastic programming (SP) to construct a model that, given a finite number of demand scenarios with discrete probability distributions, leads to an optimal delivery plan, which minimizes delivery delay penalties and supply costs for damaged buildings. A unique aspect of the model is consideration of a secondary storage warehouse (processing base) that has processing capacity and the minimum delivery unit changes as materials are processed during delivery. This prevents the waste of materials at the demand point and unnecessary deliveries, thereby reducing delivery costs and delays. However, their model does not consider the uncertainty of road conditions, which should be considered in emergency logistics. Furthermore, the parameters are not set on the actual disaster situation, which is insufficient to verify the effectiveness of the model in an actual disaster situation.

This study uses the model proposed by Oshima et al. [4] to propose the inventory routing problem (IRP) in emergency logistics that takes into account uncertainty in demand, road conditions and changes in the minimum delivery unit. We verify whether the use of secondary storage warehouses with processing capacity is effective using actual data from the 2016 Kumamoto earthquake.

2 Method for the Proposed Model

In this section, we present the assumptions of the proposed model. The objective of this model is to minimize the penalties of delivery delays, excess inventory, and delivery costs.

2.1 Notation

The following notations are used in this study.

Sets

V^d A set of one delivery base
V^c A set of n demand points
V^s A set of one secondary storage warehouse (processing base)
V^{cs} $V^c \cup V^s$ (a set of $n + 1$ nodes)
V^{dcs} $V^d \cup V^c \cup V^s$ (a set of $n + 2$ nodes, the first index (0) is the delivery base, the final index ($n + 1$) is the secondary storage warehouse (processing base))
K A set of n_k vehicles.
T A set of n_t time periods
Ξ A set of n_ξ demand scenarios $\{\xi^1, \ldots, \xi^{n_\xi}\}$

Parameters

C_{base} Maximum storage capacity at delivery base
C_i Maximum storage capacity at demand point i
C_{veh} Maximum loading capacity of vehicle
$d^{\xi}_{t,i}$ Demand at demand point i in time period t under disaster scenario ξ
$r_{i,j}$ Travel cost (travel time) between point $i \in V^{dcs}$ and point $j \in V^{dcs}$
Lot Minimum delivery unit in delivery base
$Lot^{'}$ Minimum delivery unit in secondary storage warehouse (processing base) and demand point
$RC^{\xi}_{t,i,j}$ 1 if the vehicle cannot pass between point $i \in V^{dcs}$ and point $j \in V^{dcs}$ in time period t under disaster scenario ξ, 0 otherwise
α Penalty cost due to late delivery
β Penalty cost due to excess inventory
γ Cost due to delivery time

Decision Variables

$Q^{\xi}_{t,i,k}$ Amount of deliveries from delivery base to demand point i by vehicle k in time period t under disaster scenario ξ
$x^{\xi}_{t,i,j,k}$ 1 if vehicle k passes between point $i \in V^{dcs}$ and point $j \in V^{dcs}$ in time period t in disaster scenario ξ, 0 otherwise

$y^{\xi}_{t,i,k}$ 1 if vehicle k visits demand point i in time period t under disaster scenario ξ, 0 otherwise

$z^{\xi}_{t,k}$ 1 if vehicle k is used in time period t in disaster scenario ξ, 0 otherwise

$I^{\xi}_{t,i}$ Inventory level at demand point i in time period t under disaster scenario ξ

$QC^{\xi}_{t,i,k}$ Inventory level of vehicle k shipped to point i in time period t under disaster scenario ξ

$P^{\xi}_{t,i,k}$ Amount of transshipment by vehicle k from secondary storage warehouse (processing base) to demand point i in time period t under disaster scenario ξ

2.2 Precondition

The assumptions of the proposed model are described below.

- The proposed model is represented by a complete graph connecting the delivery bases, processing bases, and demand points as vertices.
- The planning period is a multi-period.
- Delivery bases, processing bases, and demand points have inventory capacity constraints.
- A finite number of demand scenarios are denoted as $\xi \in \Xi$. The demand quantity at the demand point is given by scenario.
- In the proposed model, a single building material (lumber) is considered.
- Unfulfilled demand is carried over to the next period.
- The building material is given a minimum delivery unit constraint that varies with processing.
- The minimum delivery unit of the building material is smaller after processing than before.
- The loss of building materials due to processing is not taken into account.
- A predetermined number of vehicles are given to the delivery bases.
- All vehicles are equal and have a load capacity constraint.
- The vehicles depart from the delivery base, take at most one route per period, and return to the delivery base during that period.

2.3 Explanation of the Proposed Model

The proposed model consists of one delivery base, several demand points, and one processing base, where the processing base is located in the middle of the demand points and the delivery base (Fig. 1). The processing base is located in the middle of the demand and delivery bases. The processing base has processing and inventory storage capabilities and can store building materials sent from the delivery base without generating demand. Vehicles departing from the distribution center always pass through the processing center to supply each demand center. The minimum

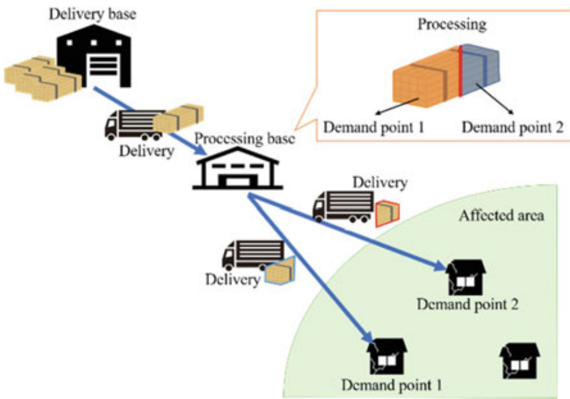


Fig. 1 The proposed model

delivery unit constraint changes as the building materials are processed at the fabrication site. When the processing base has stock, the vehicle can be loaded and delivered to the demand point as long as the capacity of the vehicle permits. In this study, existing facilities such as home centers are assumed processing bases, so installation and operation costs are not considered.

3 Formulation of the Proposed Model

In this section, we formulate the model presented in Sect. 2. The proposed model was built using the IRP model.

Minimize.

$$E_{\Xi}\left[\begin{array}{c}\sum_{t\in T}\sum_{i\in V^{cs}}\left\{\mathrm{Max}\left(-I_{t,i}^{\xi},0\right)\times\alpha+\mathrm{Max}\left(I_{t,i}^{\xi},0\right)\times\beta\right\}\\ +\sum_{i\in V^{dcs}}\sum_{j\in V^{dcs}}\sum_{t\in T}\sum_{k\in K}r_{i,j}\times x_{t,i,j,k}^{\xi}(\xi)\times\gamma\end{array}\right] \tag{1}$$

Subject to

$$I_{t,i}^{\xi}-I_{t-1,i}^{\xi}=\sum_{k\in K}(Q_{t,i,k}^{\xi}+P_{t,i,k}^{\xi})-d_{t,i}^{\xi}\quad \forall t\in T,\forall i\in V^{c},\quad \forall\xi\in\Xi \tag{2}$$

$$I_{t,i}^{\xi}\le C_i\quad \forall t\in T,\quad i\in V^{cs},\quad \forall\xi\in\Xi \tag{3}$$

$$0\le Q_{t,i,k}^{\xi}\le y_{t,i,k}^{\xi}\times C_i\quad \forall t\in T,\quad \forall i\in V^{cs},\quad \forall k\in K,\quad \forall\xi\in\Xi \tag{4}$$

$$\sum_{i\in V^{cs}}Q_{t,i,k}^{\xi}\le z_{t,k}^{\xi}\times C_{veh}\quad \forall t\in T,\quad \forall k\in K,\quad \forall\xi\in\Xi \tag{5}$$

$$\sum_{i\in V^{cs}} Q^{\xi}_{t,i,k} \equiv 0(mod\ Lot) \quad \forall t \in T, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{6}$$

$$QC^{\xi}_{t,0,k} = \sum_{i\in V^{cs}} Q^{\xi}_{t,i,k} \quad \forall t \in T, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{7}$$

$$QC^{\xi}_{t,i,k} = \sum_{n\in V^{dcs}} \left(x^{\xi}_{t,n,i,k} \times QC^{\xi}_{t,n,k}\right) - Q^{\xi}_{t,i,k} + P^{\xi}_{t,i,k} \quad \forall t \in T, \quad \forall i \in V^{cs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{8}$$

$$QC^{\xi}_{t,0,k} \le C_{veh} \quad \forall t \in T, \forall i \in V^{cs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{9}$$

$$\sum_{i\in V^{cs}} \sum_{k\in K} Q^{\xi}_{t,i,k} \le C_{base} \quad \forall t \in T, \quad \forall \xi \in \Xi \tag{10}$$

$$\sum_{i\in V^{cs}} x^{\xi}_{t,i,j,k} = \sum_{i\in V^{cs}} x^{\xi}_{t,j,i,k} = y^{\xi}_{t,j,k} \quad \forall t \in T, \quad \forall j \in V^{cs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{11}$$

$$\sum_{i\in V^{cs}} x^{\xi}_{t,0,i,k} = \sum_{i\in V^{cs}} x^{\xi}_{t,i,0,k} = z^{\xi}_{t,k} \quad \forall t \in T, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{12}$$

$$y^{\xi}_{t,i,k} \le z^{\xi}_{t,k} \forall t \in T, \forall i \in V^{cs}, \forall k \in K, \forall \xi \in \Xi \tag{13}$$

$$\sum_{i\in V^{dcs}} \sum_{j\in V^{dcs}} x^{\xi}_{t,i,j,k} = 0 \quad \forall t \in T, \quad \forall k \in K, \quad i = j, \quad \forall \xi \in \Xi \tag{14}$$

$$\sum_{i\in S} \sum_{j\in S} x^{\xi}_{t,i,j,k} \le |S| - 1 \quad \forall S \subset V^{dcs}, \quad \forall t \in T, \quad \forall k \in K, \quad \forall \xi \in \Xi \quad (S \in \{1, \ldots, n\}, \quad 2 \le |S| \le n - 1). \tag{15}$$

$$x^{\xi}_{t,i,j,k}, y^{\xi}_{t,i,k}, z^{\xi}_{t,k} \in \{0, 1\} \quad \forall t \in T, \quad \forall i \in V^{dcs}, \quad \forall j \in V^{dcs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{16}$$

$$I^{\xi}_{t,n+1} = I^{\xi}_{t-1,n+1} + \sum_{k\in K} Q^{\xi}_{t,n+1,k} - \sum_{i\in V^{cs}} \sum_{k\in K} P^{\xi}_{t,i,k} \quad \forall t \in T, \quad \forall \xi \in \Xi \tag{17}$$

$$Q^{\xi}_{t,i,k} \equiv 0\left(mod\ Lot'\right) \quad \forall t \in T, \quad \forall i \in V^{cs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{18}$$

$$x^{\xi}_{t,0,n+1,k} = z^{\xi}_{t,k} \quad \forall t \in T, \forall k \in K, \quad \forall \xi \in \Xi \tag{19}$$

$$P^{\xi}_{t,i,k} \le y^{\xi}_{t,i,k} \times C_i \quad \forall t \in T, \quad \forall i \in V^{cs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{20}$$

$$0 \leq \sum_{i \in V^c} \sum_{k \in K} P^{\xi}_{t,i,k} \leq I^{\xi}_{t-1,n+1} \quad \forall t \in T, \quad \forall \xi \in \Xi \tag{21}$$

$$P^{\xi}_{t,i,k} \equiv 0 \big(mod\ Lot' \big) \quad \forall t \in T, \quad \forall i \in V^{cs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{22}$$

$$P^{\xi}_{t,n+1,k} = 0 \quad \forall t \in T, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{23}$$

$$x^{\xi}_{t,i,k,k} \times RC^{\xi}_{t,i,j} = 0 \quad \forall t \in T, \quad \forall i \in V^{dcs}, \quad \forall j \in V^{dcs}, \quad \forall k \in K, \quad \forall \xi \in \Xi \tag{24}$$

Objective function (1) minimizes the average value of the sum of the delivery delay penalty, excess inventory penalty, and delivery cost of the building material in all demand scenarios. The left-hand side of constraint (2) represents the change in inventory quantity between the end of period t and the end of period t − 1, and the right-hand side represents the difference between the delivery quantity and the demand quantity in period t. Constraint (3) guarantees that the inventory quantity fits into the inventory capacity. Constraint (4) indicates that the number of deliveries by vehicle does not exceed the inventory capacity. Constraint (5) guarantees the loading capacity constraint of the vehicle. Constraint Eq. (6) represents the minimum delivery unit constraint from the delivery base to the processing base. Constraint (7) shows the amount of inventory when the vehicle leaves the delivery base. Constraints (8) and (9) ensure that the inventory quantity of the vehicle obeys the loading constraint when the vehicle leaves the demand point. Constraint (10) guarantees that the total number of building materials delivered from the delivery base does not exceed the capacity of the delivery base. Constraint (11) ensures the preservation of vehicle flow at the demand point. Constraint (12) guarantees that only vehicles making deliveries can leave the delivery base and that all vehicles must return to the delivery base. Constraint (13) ensures that vehicles that do not leave the delivery base do not visit the demand point. Constraint (14) prevents vehicles from traveling between the same points. Constraint (15) avoids the occurrence of the partial circuits. Constraint (16) defines the variables. Constraint (17) represents the inventory storage equation at the processing base. Constraint (18) shows that the minimum delivery unit constraint after processing is performed. Constraint (19) ensures that the vehicle leaving the delivery base always visits the processing base first. Constraint (20) guarantees that the building materials delivered from the processing base do not exceed the capacity of the demand point. Constraint (21) ensures that the number of deliveries from the processing base to the demand point does not exceed the inventory of the processing base in the previous period. Constraint (22) represents the minimum delivery unit constraint for building materials delivered from the processing base. Constraint (23) prevents deliveries from the processing base to the processing base. Constraint (24) represents the road condition constraint.

Table 1 Parameter

n_d	n_c	n_s	n_k	n_t	n_ξ		
1	4	1	2	6	20		
c_{base}	c_i	c_{veh}	Lot	Lot'	α	β	γ
900	300	500	20	1	10	1	0.5

4 The Case Description

In this section, we investigate the effectiveness of the proposed model using numerical experiments. We conducted numerical experiments using five case studies with 20 demand scenarios per case to examine the effect of considering processing locations with and without constraints on road conditions.

4.1 Numerical Example

The numerical example is given (Table 1).

In this study, the model is applied to several cities that were severely damaged by the Kumamoto earthquake in 2016. The travel cost in the model was set by calculating the travel time by car between each city using ArcGIS (Fig. 2). In Fig. 2, the darker areas represent the areas closer to the epicenter with higher seismic intensity, and the numbers on the lines represent the distance (km) of the lines. The demand was set using a normal distribution with randomness, referring to actual data on housing restoration after the earthquake. The probability of the road condition becoming impassable was determined by the seismic intensity, and the probability was determined using the higher seismic intensity of the two connected nodes. We assumed roads were to be repaired, and the probability of impassability decreased with time. The average of all scenarios shows that the maximum value of the impassable edge is 37% in the first period and 5% in the last period.

The calculations were performed using Windows 10 home 64 bit with Intel® Core ™ i9-9900X CPU @ 3.50 CPU and 128 GB RAM on Python 3.8.5. The mathematical programming solver used was the Gurobi Optimizer version 9.1.2.

4.2 Computational Results

The results were compared for the four models depending on whether the processing base and road conditions were considered (Fig. 3). For each scenario, the model on the left is the model that does not consider both the processing base and road conditions (IRP), followed by the model that only takes into account the processing base (IRPF); next to it is the model that only considers road conditions (IRPR); and

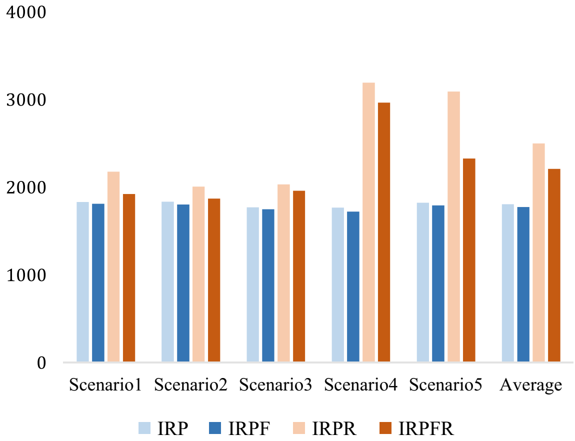


Fig. 2 Application area of the proposed model

finally the model that takes into account both (IRPFR). In all the case studies, the IRPF model recorded slightly better objective function values than the IRP model. Comparing the IRP model with the IRPR model, the IRPR model is unable to pass through roads (edges), which increases the cost of travel or increases the penalty for late delivery, and increases the objective function value in all case studies. Comparing the IRPR model with the IRPFR model, the IRPFR model shows better results in all case studies, and the improvement is larger than that of the IRP and IRPF models. This result suggests that it may be more reasonable to set up a processing base at an intermediate location for delivery after an earthquake when roads are unavailable.

5 Conclusion

We proposed an inventory routing planning model for emergency logistics for the early recovery of damaged buildings in this study. The feature of this study is that delivery of building materials status is changed by stopping at processing bases during the delivery process, taking into consideration road conditions. The proposed model considers the uncertainty of the demand scale and road network, as well as the change in the processing state of the building materials. The former is modeled using stochastic programming with a finite number of scenarios. The latter was incorporated into the model by assuming that the constraint on the minimum delivery unit changed during the delivery process. The model proposed by Oshima et al. (2020) [4], which considers the processing base (IRPF), is slightly better than the previous model (IRP) in all scenarios, even when using data from actual earthquake damage. In addition, the proposed model (IRPFR) is better than the model without considering the processing bases (IRPR). This improvement rate is larger than the case without considering the road conditions, which confirms the effectiveness of using processing

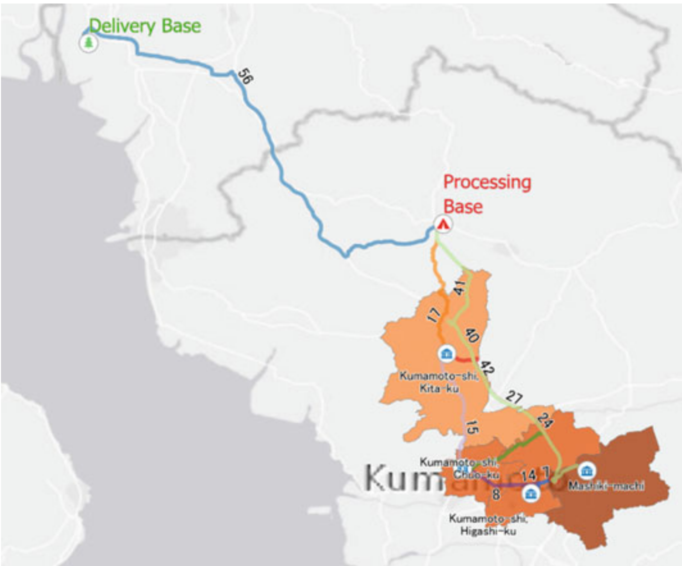


Fig. 3 Comparison of the objective function values of each model in the scenarios

bases for delivery in the event of an earthquake when part of the road network is not available.

Future work includes optimization of the number and location of processing bases, and optimization using meta-heuristic algorithms on a larger scale.

References

1. Yuan, Y., Wang, D.: Path selection model and algorithm for emergency logistics management. Comput. Ind. Eng. **56**(3), 1081–1094 (2009)
2. Tofighi, S., Torabi, S.A., Mansouri, S.A.: Humanitarian logistics network design under mixed uncertainty. Eur. J. Oper. Res. **250**(1), 239–250 (2016)
3. Rawls, C.G., Turnquist, M.A.: Pre-positioning of emergency supplies for disaster response. Transp. Res. Part B: Methodol. **44**(4), 521–534 (2010)
4. Tomoki, O., Aya, I., Takumi, I., Kenjiro, M.: Modeling of inventory routing problem considering collaborative strategy and change of minimum delivery unit constraint in emergency logistics. In: 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), vol. 1, pp. 729–734 (2020). https://doi.org/10.1109/IIAI-AAI50415.2020.00146

Problem of Modeling Global and Closed-Loop Supply Chain Network Design

Hiromasa Ijuin and Tetsuo Yamada

Abstract Nowadays, the economic activities cause some environmental problems for not only the global warming by Greenhouse Gas (GHG), but also the material starvation by mass consumption of resources. In order to resolve their problems simultaneously, Japanese government took effect with the plastic regulation and announced that GHG emissions became zero by 2050. In order for manufacturing companies to develop sustainably, they are required to design the Global Closed-Loop Supply Chain (GCLSC) network to recycle End-of-Life (EOL) products and reduce GHG emissions. The GCLSC network integrates the global supply chain network, which is a series of cross-border transaction of product, and the local reverse supply chain network, which is a domestic transaction for collecting and recycling EOL products. However, it is necessary to select appropriate global suppliers considering GHG emissions because the amount of material-based GHG emissions in EOL product is dependent on manufacturing country. On the other hand, the GCLSC network is costly for procurement, manufacturing, transportation, recycling and opening facilities costs. Thus, the decision maker who designs designing GCLSC network should consider not only cost, but also environmental aspects such as recycling EOL product and GHG saving weight simultaneously. This study designs the GCLSC network to minimize total cost and to maximize entire recycling rate on global and local supply chain network. Furthermore, the GHG emission from components procurement and the GHG saving weight by recycling EOL product is evaluated. First, the GCLSC network is modeled. Next, the objective function for minimizing total cost and maximizing entire recycling rate are formulated with integer programming. Finally, a numerical experiment is conducted and evaluated in terms of costs, entire recycling rate, and GHG emission.

Keywords Global supply chain · Reverse logistics · Lifecycle inventory database · BOM (bill of materials) · 0–1 integer programming

H. Ijuin · T. Yamada (✉)
The University of Electro-Communications, Tokyo, Japan
e-mail: tyamada@uec.ac.jp

H. Ijuin
e-mail: hju1n@yahoo.co.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_6

1 Introduction

Nowadays, the assembled product, such as vacuum cleaner and smartphone is delivered to users through a global supply chain that spans procurement, manufacturing, and sales in multiple countries [1]. The economic activities such as those have caused the Marine-Plastic-Issue and global warming [2]. The Marine-Plastic-Issue is caused by disposing End-of-Life (EOL) product made from plastic materials, and global warming is occurred by Greenhouse Gas (GHG) emission in manufacturing. In order to resolve their problems, governments and manufacturing companies are taking action in around the world. For example, EU adopt the low “the reduction of impact of certain plastic products on the environment” in 2018 and decided to ban the single-used plastic products, which was not recycled, such as straw and cutlery by 2021 [3]. The Coca-Cola Company announced that they would collect as much cans and plastic bottles as the world’s product sales volume by 2030 [4]. Japanese government took effect with the plastic regulation and announced that GHG emissions became zero by 2050 [2]. In order for manufacturing companies to develop sustainability, they are required to design the Global Closed-Loop Supply Chain (GCLSC) network for reducing GHG emissions and recycling EOL products in manufacturing stages [5].

The GCLSC integrates the global supply chain network [6], which is a series of cross-border transaction of products, and the local reverse supply chain network [7], which is a local transaction for collecting and recycling EOL products because as accordance with the Basel Convention [8], the collection center, the recovery facility, and the disposal facility must be located in the same country as the place of market. However, GCLSC network is costly for procurement, manufacturing, transportation, recycling and opening facilities costs. Thus, the decision maker who designs configuration of GCLSC network should consider to select appropriate global suppliers considering material-based GHG emissions because the amount of material-based GHG emissions in products is dependent on power generation methods such as thermal power generation, hydroelectricity, and solar photovoltaics in each country [6]. As a previous study, Armin et al. [9] proposed CLSC model to determine facility location decisions and lateral transshipment quantities that minimize the total supply chain cost across different disruption scenarios. However, they did not evaluate the environmental impacts such as recycling rate and GHG emissions. Aldoukhi and Gupta [10] designed CLSC considering to minimize total cost and carbon emission and to maximize service level of retailers. However, they did not consider materials such as plastic and aluminum obtained from recycling EOL product. Polat et al. [11] integrated forward and reverse logistics network design. However, they did not take into account that different countries have different costs and GHG emissions when procuring and manufacturing components.

This study designs the GCLSC network to (1) minimize total cost and (2) maximize the entire recycling rate by Integer Programming (IP) [12] with ε constraint for entire recycling rate on global supply chain network and domestic closed-loop, where the GHG emission from components procurement and the GHG saving weight by recycling EOL product is calculated. First, the GCLSC network is draw and modeled. Next, the objective functions for minimizing total cost and the environmental evaluation for entire recycling rate and GHG emission are formulated with IP [12]. Finally, a numerical experiment is conducted and evaluated in terms of costs, entire recycling rate, and material-based GHG emission weight.

2 Model

2.1 Assumption and Notation

Figure 1 shows a GCLSC network model proposed in this study. That integrates a global supply chain network [6] and a local reverse supply chain network [7]. In the global supply chain network, the component j is procured from supplier o and transported to factory p for manufacturing an assembly product. After that, the assembly product is transported from factory p to market q. In the market q, assembly product is sold to an user and returned by the user as the EOL product. When component j is procured from supplier o, the material-based GHG emission GHG_{oj} is generated, and the amount of it is depends on the country where supplier o is in.

In the local reverse supply chain network, EOL products are transported from market q to collection center g. In a collection center g, EOL products are identified and divided status s based on user situation. Furthermore, EOL product with status s is transported from collection center g to recovery facility r or disposal facility h.

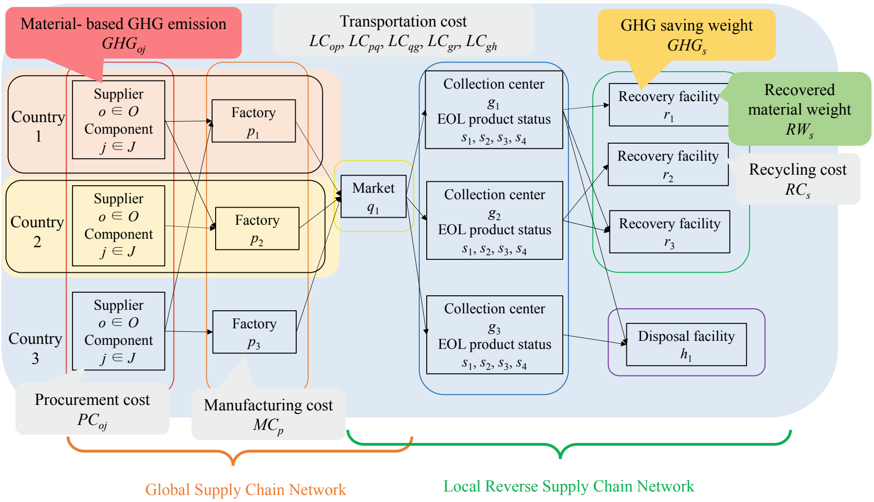


Fig. 1 Global closed-loop supply chain network model: Case of 3 countries

When the EOL product is transported to recovery facility, it is recycled based on EOL product status *s*. On the other hand, when the EOL product is transported to disposal facility, it is disposed.

In this study, material-based GHG emission by each material GHG_{oj} for component *j* is different based on procured country from supplier *o*. Moreover, recycling cost RC_s, recovered material weight RW_s, and GHG saving weight GHG_s and by recycling EOL product is dependent on the status *s*. The recycling cost RC_s includes disassembly, processing and disposal costs. The disassembly cost is based on disassembly time by Recyclability Evaluation Method (REM) [13]. The processing cost relates to shredding and so on. The disposal cost is the sum of final processing cost such as landfill and sale of recycling material. The recovered material weight RW_s means the recovered material weight by recycling EOL product.

The notations used in this study are as follows:

Sets:

- O Set of suppliers, $o \in$ O
- J Set of components, $j \in J$
- P Set of factories, $p \in P$
- Q Set of markets, $q \in$ Q
- G Set of collection centers, $g \in G$
- R Set of recovery facilities, $r \in R$
- H Set of disposal facilities, $h \in H$
- S Set of statues of EOL product, $s \in$ S

Variables:

- v_{opj} Amount of components *j* transported from supplier *o* to factory *p*
- v_{pq} Amount of products transported from factory *p* to market *q*
- v_{qg} Amount of EOL products transported from market *q* to collection center *g*
- v_{grs} Amount of EOL products with status *s* transported from collection center *g* to recovery facility *r*
- v_{ghs} Amount of EOL products with status *s* transported from collection center *g* to disposal facility *h*
- u_p 1 if the fixed opening cost at factory *p* is opened, and 0 otherwise
- u_r 1 if the fixed opening cost at recovery facility *r* is opened, and 0 otherwise
- TC Total cost [US$]
- TGC Total cost of global supply chain network [US$]
- TRC Total cost of local reverse supply chain network [US$]
- RR Entire recycling rate of local reverse supply chain network [%]
- ε_{RR} Epsilon rate of entire recycling rate [%]
- $TGHG$ Total GHG emission on global GCLSC [g-CO_2eq]
- GHG_{oj} GHG emission weight for component *j* procured from supplier *o* [g-CO_2eq]
- GHG_s GHG saving weight by recycling EOL product with status *s* [g-CO_2eq]

Parameters:

LC_{op} Transportation cost of a component from supplier *o* to factory *p* [US$]
LC_{pq} Transportation cost of a product from factory *p* to market *q* [US$]
LC_{qg} Transportation cost of a EOL product from market *q* to collection center *g* [US$]
LC_{gr} Transportation cost of a EOL product collection center *g* to recovery facility *r* [US$]
LC_{gh} Transportation cost of a EOL product collection center *g* to disposal facility *h* [US$]
PC_{oj} Procurement cost of component *j* by supplier *o* [US$]
MC_{op} Manufacturing cost per product at factory *p* [US$]
RC_s Recycling cost per EOL product with status *s* [US$]
FC_p Fixed cost for opening factory *p* [US$]
FC_r Fixed cost for opening recovery facility *r* [US$]
RW_s Recovered material weight of EOL product with status *s* [g]
PW Total weight of a EOL product [g]

2.2 Formulation

This GCLSC model has a bi-objective problem for minimizing total cost *TC* and for maximizing the entire recycling rate *RR*. Regarding the total cost *TC* on GCLSC network, Eq. (1) minimizes of total cost of global supply chain network *TGC* and local reverse supply chain network *TRC*. Equation (2) maximizes the entire recycling rate *RR* on GCLSC network which is the ratio of total recovered material weight of recycling EOL product at recovery facilities against the total collected material weight of collecting EOL product at collection centers. Equation (3) means the total cost of global supply chain *TGC* including the procurement cost PC_{oj} for component *j* from supplier *o*, the manufacturing cost MC_p of product in factory *p*, the opening factory cost FC_p for assembly factory *p*, and the transportation cost LC_{op} and LC_{pq} transported from supplier *o* to factory *p* and from factory *p* to market *q*. Equation (4) represents the total cost of local reverse supply chain network *TRC* including the recycling cost *RCs* for recycling EOL product with status *s*, the opening facility cost FC_r for recovery facility *r*, and the transportation cost LC_{qg}, LC_{gr}, and LC_{gh} transported from market *q* to collection center *g* and from collection center *g* to recovery facility *r* or disposal facility *h*. Equation (5) is epsilon constraint of entire recycling rate *RR* for GCLSC network. Equation (6) shows the total GHG emission *TGHG* which is the difference between the sum of GHG emission weight GHG_{oj} for procured component *j* minus the sum of GHG saving weight GHG_s for recycled EOL product with status *s*.

Objective functions and GHG evaluation:

$$TC = TGC + TRC \rightarrow \min \tag{1}$$

$$RR = \frac{\sum_{g\in G}^{G}\sum_{r\in R}^{R}\sum_{s\in S}^{S} RW_s v_{grs}}{\sum_{q\in Q}^{Q}\sum_{g\in G}^{G} TW v_{pq}} \to \max \tag{2}$$

$$TGC = \sum_{o\in O}^{O}\sum_{p\in P}^{P}\sum_{j\in J}^{J}(PC_{oj} + LC_{op})v_{opj} + \sum_{p\in P}^{P}\sum_{q\in Q}^{Q}\left(LC_{pq} + MC_p\right)v_{pq} + \sum_{p\in P}^{P} FC_p u_p \tag{3}$$

$$TRC = \sum_{q\in Q}^{Q}\sum_{g\in G}^{G} LC_{qg}v_{qh} + \sum_{g\in G}^{G}\sum_{r\in R}^{R}\sum_{s\in S}^{S}\left(RC_s + LC_{gr}\right)v_{grs} + \sum_{g\in G}^{G}\sum_{h\in H}^{H}\sum_{s\in S}^{S} LC_{gr}v_{grh} + \sum_{r}^{R} FC_r u_r \tag{4}$$

$$RR \geqq \varepsilon_{RR} \tag{5}$$

$$TGHG = \sum_{o\in O}^{O}\sum_{p\in P}^{P}\sum_{j\in J}^{J} GHG_{oj}v_{opj} - \sum_{g\in G}^{G}\sum_{r\in R}^{R}\sum_{s\in S}^{S} GHG_s v_{grs} \tag{6}$$

3 Problem Example

To design the GCLSC network, a problem example is prepared to integrate a global supply chain and local reverse supply chain. A vacuum cleaner with 23 types of components is used as examples of assembled and EOL product as well as Hayashi et al. [14], Ijuin et al. [7], and Ishii et al. [15]. It is assumed that a motor in vacuum cleaner is excluded because the procurement cost of the motor is about 95% of the sum of procured components for all the components. The details of GCLSC example are prepared in this study as follows.

- As suppliers, 52 cities are set from 4 country for China, Malaysia, the United States, and Japan, in other word, each country has 13 suppliers.
- As candidate factories' location, 4 cities for Shanghai, Kuala Lumpur, Seattle, and Tokyo, are selected. The production capacity of each factory is 6,000 units, and the market demand set at Tokyo, the Japan, is 12,000 units.
- As collection centers, 3 sites in Itabashi, Koto, and Ota where are in Japan, are prepared.

- As recovery facilities for recycling EOL product, 3 sites in Ota, Chiba, and Chofu where are in Japan, are selected. They have recyclable capability based on status *s* of EOL product.
- As disposal facility, Koto site where is in Japan, is prepared.

The statuses of EOL products in the case of vacuum cleaner are also prepared based on disassembly scenarios [13]. From disassembly scenarios, the recycling cost RC_s and recovered material weight RW_s, and GHG saving weight GHG_s are divided 4 statuses. For example, the EOL product with status number #4 has 1.53 [US$] for recycling cost, 952.79 [g] recovered material weight, and 1276.94 [g-CO_2eq] GHG saving weight by recycled components in EOL product and it is recycled in recovery facility in Chiba or Ota.

4 Result

To solve the bi-objective problem for minimizing total cost and maximizing the entire recycling rate, the entire recycling rates *RR* between from 0 [%] (minimum) to 60 [%] (maximum) are changed by epsilon constraint ε_{RR}. Figure 2 shows the total cost and the total material-based GHG emission of GCLSC network with the entire recycling rate *RR*. The sum of manufacturing cost and procurement cost was the highest, accounting for over 90% of the total cost. Even if any target recycling rate were set, the total transportation cost on global supply chain, the total procurement cost, the total manufacturing cost, and total opening facility cost for assembly factories were not changed in the all cases. Because of that, the entire recycling rate can not impact on the global supply chain in the example. On the other hand, the total transportation cost on the local reverse supply chain, the total recycling cost and the total opening facility cost were increased. One of the reasons was that lager amount of EOL products was recycled in attempting to increase the entire recycling rate.

From Fig. 2, when the entire recycling rate was increased by 20 [%], the total GHG emission was decreased by 6,000 [kg-CO_2eq]. This is because weight of recovered materials was less GHG emission than procuring virgin one. Therefore, it is found that GHG emissions could be lower at the same time as increasing the recycling rate.

Figure 3 shows the GCLSC network in the case with the maximal entire recycling rate (60 [%]). In the case, Chinese and Malaysian suppliers were selected. Additionally, from Fig. 2, 92% of the total cost was occupied by total manufacturing cost and total procurement cost. Therefore, candidate locations with lower procurement and manufacturing cost, such as suppliers and factories in China and Malaysia, were selected.

On the other hand, Chiba's recovery facility in Japan whose opening cost was the highest in the candidate locations is opened. One of the reasons is that was the recovery facilities in Ota and Chofu can not recycle the EOL products with statuses 1 (All components disassembled) and 2 (Recycling rate maximum) assumed as in this

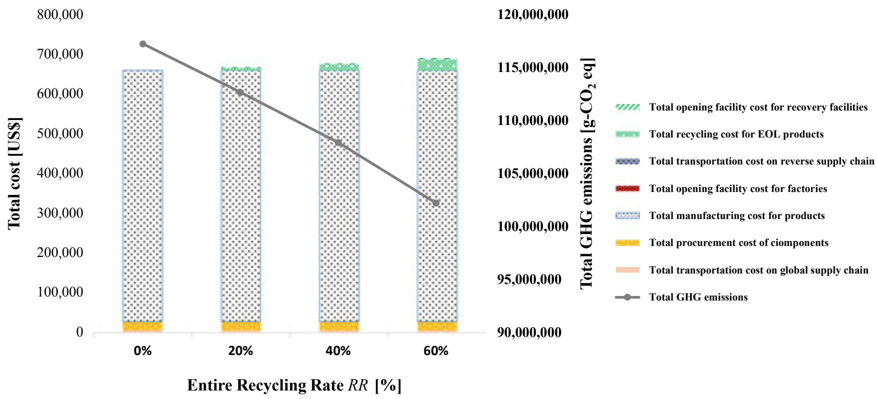


Fig. 2 Total cost and total GHG emission of GCLSC with entire recycling rate

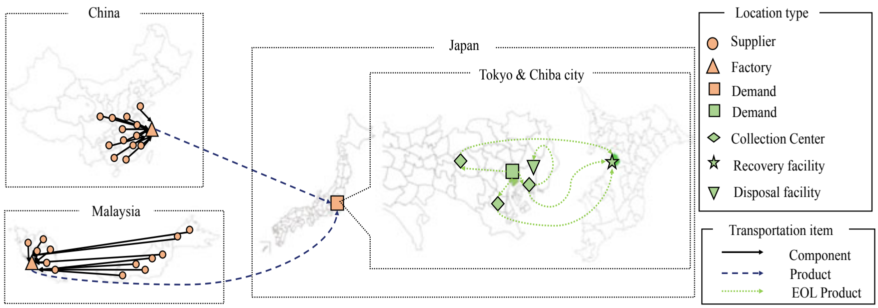


Fig. 3 GCLSC network flow of component, product, and EOL products with the maximal entire recycling rate (60 [%]) (Map source: https://www.freemap.jp/)

study. In order to increase the entire recycling rate, all EOL products were transported and recycled in Chiba's recovery facility in the case.

5 Conclusion

This study designed the GCLSC network and observed changes of the costs, entire recycling rate, total GHG emission, and network flow for component, product and EOL product.

In the GCLSC network, costs in global forward supply chain such as component procurement and manufacturing product was by 9 times higher than ones in the local reverse supply chain. Thus, total GHG emission was decreased because larger amount of EOL products were recycled when the entire recycling rate is increased. In the global supply chain flow, suppliers and factory were selected in China and

Malaysia for saving procurement cost and manufacturing cost. On the other hand, as a local reverse supply chain flow, Chiba's recovery facility which had the highest opening was opened in order to increase the entire recycling rate.

Further studies should consider a multi-objective model to minimizing cost, to maximize recycling rate, and to minimize GHG emission simultaneously.

Acknowledgements This research was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Grant-in-Aid for Scientific Research (A), JP18H03824, from 2018 to 2023.

References

1. Ravindran, A.R., Warsing, D.P., Jr.: Supply Chain Engineering, CRC Press (2013)
2. Ministry of the Environment: Annual report on the environment: the sound material-cycled society and biodiversity in Japan 2020, NiKKEI PRINTING INC (2020). (In Japanese)
3. The European Parliament and Council: The reduction of the impact of certain plastic products on the environment (2019). https://data.consilium.europa.eu/doc/document/PE-11-2019-INIT/en/pdf (Accessed on 8 Jul, 2021).
4. Gendatsu, K.: Challenge to deplasticization. Yama-kei Publishers (2020). (In Japanese)
5. Santander, P., Cruz, F., Boudaoud, H., Camargo.: Closed loop supply chain network for local and distributed plastic recycling for 3D printing: a MILP-based optimization approach. Resour. Conserv. Recycl. **154**, 104531 (2019)
6. Urata, T., Yamada, T., Itubo, N., Inoue, M.: Global supply chain network design and Asian analysis with material-based carbon emissions and tax. Comput. Ind. Eng. **113**, 779–792 (2017)
7. Ijuin, H., Kinoshita, Y., Yamada, T., Ishigaki, A., Inoue, M.: Designing reverse supply chain network with costs and recycling rate by using linear physical programming. Int. J. Smart Comput. Artif. Intell. **3**(2), 57–76 (2019)
8. Kei., O.: Addressing plastics under the "basel convention on the control of transboundary movements of hazardous wastes and their disposal" and other international developments. Jap. Soc. Mater. Cycles Waste Manag. **30**(2), 98–105 (2019). (In Japanese)
9. Armin, J., Michael, H., Amir, K.: Closed-loop supply chain network design under disruption risks: a robust approach with real world application. Comput. Ind. Eng. **116**, 178–191 (2017)
10. Aldoukhi, M., Gupta, S.M.: Use of maximal covering location problem to design a closed loop supply chain network under product substitution. Appl. Manag. Sci. **20**, 71–96 (2020)
11. Polat, O., Ünal, Ö.F., Plot, L.Ö., Güngör, A.: Integrated forward-reverse logistics network design: an application for electrolytic copper conductor reel distribution. In: Alqahtani , A.Y., Kongar, E., Pochampally, K.K., Gupta, S.M. (Editors), Responsible Manufacturing. CRC Press. 85–103 (2019)
12. Hillier, F.S., Lieberman, G.J.: Introduction to Operations Research, 8th edn. McGraw-Hill (2005)
13. Igarashi, K., Yamada, T., Inoue, M.: 2-Stage optimal design and analysis for disassembly system with environmental and economic parts selection using the recyclability evaluation method. Int. J. Ind. Eng. Manag. Syst. **13**(1), 52–66 (2014)
14. Hayashi, K., Matsumoto, R., Yamada, T., Nagasawa, K., Kinoshita, Y.: Designing a global supply chain network considering carbon tax and trans-pacific partnership agreement. In: Proceedings of the 2020 The Society of Plant Engineering Japan Spring Meeting, 97–100 (2020). (In Japanese)
15. Ishii, M., Nakamura, K., Yamada, T., Inoue, M.: Design of closed-loop supply chain network considering return and recycling rate. J. Soc. Plant Eng. Jpn **31**(1), 1–13 (2019). (In Japanese)

Vehicle Relocation Scheduling Considering Charging Time for One-Way Electric Vehicle Car-Sharing Systems

Aya Ishigaki, Akane Hagimoto, and Tomoki Sanada

Abstract Traditional car-sharing services are based on a two-way scheme, wherein the user picks up and returns a vehicle at the same parking station. Some services also permit one-way trips; that is, the user is allowed to return the vehicle to another parking station. The one-way scheme is more widely available for users but may lead to an imbalance in the number of vehicles at each parking station; in other words, the user demand in each station may not be satisfied. In such cases, the service provider can relocate the vehicles to create a better distribution among the parking stations. In the case of electric vehicle (EV) car-sharing, such a problem is more complex because the travel range depends on the charge level of the battery. Moreover, EV relocation can lead to an imbalance in staff members between stations. Thus, staff members themselves need to be relocated between stations to perform vehicle relocations, considering the waiting time for charging the EV. In this study, we consider the charging and relocation problems of an EV car-sharing system. Additionally, we demonstrate the joint optimization of vehicle relocation and staff scheduling that can minimize the waiting time for staff and EVs at parking stations.

Keywords Electric vehicle · One-way car-sharing · Vehicle relocation problem · Staff scheduling · Waiting time

1 Introduction

1.1 Environmental Impact of Automobiles

The environmental impact of automobiles is a serious issue. Therefore, the Japanese government has declared a ban on the sale of gasoline-powered vehicles by 2030 and has begun to promote a switch to electric vehicles (EVs) and fuel cell vehicles. However, the CO_2 emissions generated during the manufacture of EVs are said to

A. Ishigaki (✉) · A. Hagimoto · T. Sanada
Tokyo University of Science, Chiba, Japan
e-mail: ishigaki@rs.tus.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_7

account for more than half of the total lifecycle of vehicles, and there is a great need to address this issue [1]. In addition, owing to the impact of COVID-19, many countries have implemented a “social distancing strategy” to reduce opportunities for human contact. Consequently, the number of people using public transportation has decreased significantly, whereas the number of people using cars and bicycles has increased.

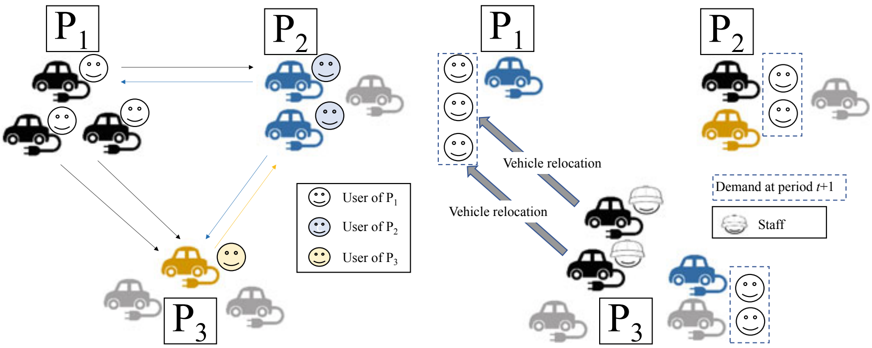
1.2 Car-Sharing Systems

Car-sharing is effective in reducing the environmental impact of manufacturing because it can improve the utilization rate of cars and reduce the number of cars owned by multiple users. It is a system in which cars are rented out to members who have registered as users in advance; they can use the cars as and when required. In recent years, the demand for car-sharing has been increasing, particularly in Europe, and many companies such as Daimler in Germany (which operates Car2Go) and Renault in France provide car-sharing systems [2, 3].

There are two types of car-sharing operations: one-way and round-trip (two-way). One-way use allows a user to return a car at a different location from the departure point, whereas in round-trip use, users will have to return the vehicle at the departure point. The one-way car-sharing system has been adopted by many businesses because it is more flexible and convenient for users than the round-trip system. However, it requires the relocation of vehicles by the operator’s staff to solve the resulting uneven distribution of vehicles. Jorge et al. [4] developed a linear model to determine optimal staff relocation. In the case of EVs, it is necessary to manage their remaining battery power, and they can be recharged in the middle of the night when not in use. However, in the case of car-sharing systems, users need to stop by the charging spot to recharge the battery or recharge the battery after the car has been returned to the parking station. In other words, in a one-way car-sharing system, it is necessary to consider not only the relocation of staff to eliminate the uneven distribution of vehicles, but also the charging schedule. Although there are some studies on one-way car-sharing systems that consider the schedules of staff for relocation, none of them consider the charging schedules of EVs.

1.3 Purpose of This Study

This study aims to develop a model for the charging schedule and staff relocation in car-sharing systems for EVs. Additionally, it aims to simultaneously optimize the vehicle repositioning plan and the charging schedule to minimize the waiting time of the staff at the station and that of the EV.



(a) Initial location of electric vehicles (EVs) at period *t*. (b) Final location of EVs at period *t*.

Fig. 1 Image of relocation operation of EVs in a one-way car-sharing system

2 Model

2.1 *Vehicle Relocation Scheduling*

Figure 1 illustrates the relocation of EVs in a one-way car-sharing system. Table 1 presents the number of EVs moving from parking station P_i to parking station P_j in period t and the number of EVs required in period $t + 1$. In period t, there are three, two, and one EVs at parking stations P_1, P_2, and P_3 respectively, and each EV is assigned to a user and destination. Once the EVs move, there are one, two, and three EVs at parking stations P_1, P_2, and P_3, respectively. Because the demand at each parking station P_1, P_2, and P_3 in period $t + 1$ is three, two, and one, respectively, there is a shortage of two vehicles at P_1 and a surplus of two vehicles at P_3. Therefore, a staff member visits parking station P_3 and moves two EVs to parking station P_1 until the beginning of the period $t + 1$.

Jorge et al. [4] developed a new mathematical model to optimize vehicle repositioning operations in a one-way car-sharing system. The objective was to maximize the revenue, including user payments, repositioning costs, vehicle fuel costs, vehicle cost depreciation, and park station installation costs. The results indicated that redeployment had the potential to significantly increase revenue. However, only the additional vehicle fuel required and staffing costs due to relocation were considered, and not the time required for vehicle relocation.

2.2 *Charging Schedule for Electric Vehicles*

There are two types of charging facilities for EVs, namely ordinary and quick. Ordinary chargers are of two types, with voltages of 200 and 100 V, and a full charge

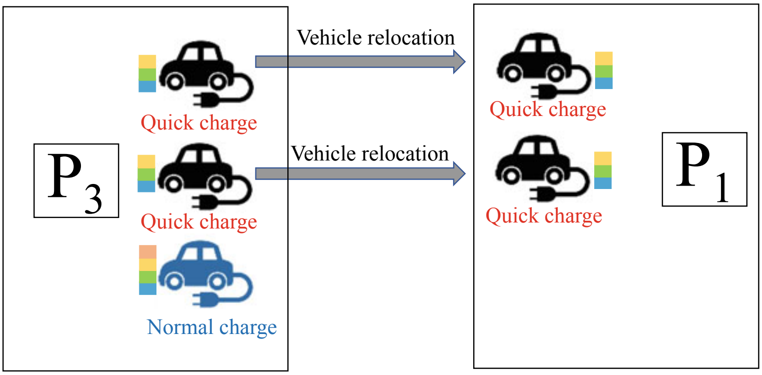


Fig. 2 Image of application of selection rules for EVs to be relocated in a one-way car-sharing system

Table 1 Details of an electric vehicle (EV) moving between parking stations

		To			Demand at period (t+1)
		P_1	P_2	P_3	
From	P_1	0	1	2	3
	P_2	1	0	1	2
	P_3	0	1	0	2

can be achieved in 5–8 h at 200 V. Ordinary chargers are suitable for use in places where cars are parked for long periods of time, such as residences, offices, and lodging facilities because the cost of installing the equipment is low. In contrast, quick chargers can produce a higher voltage and can charge up to approximately 80% of the vehicle's total capacity in approximately 30 min [5]. Because their charging speeds are suppressed for battery protection when the rate exceeds approximately 80%, these chargers are suitable for charging on the go or for emergency charging. In the case of a one-way car-sharing system, it is difficult to charge for a long period of time the EVs that are to be relocated, because they need to be moved between the end of period t and the beginning of period $t + 1$. In this study, we selected the EVs to be relocated from among the ones with the largest remaining battery capacities and moved them to the next parking station and recharged them using a quick charger. In other words, it is assumed that 100% of the battery can be recharged by normal charging and 75% of the remaining battery capacity at the beginning of a period can be recharged after relocation by quick charging, considering the battery drain (Fig. 2).

2.3 Mathematical Model

In this study, we modified the mathematical model of Jorge et al. [4] to consider the effect of charging time on EVs. The main notations and objective functions are as follows:

[notation]

$R_{i_t,j_{t+\delta_{ij}^t}}$	number of vehicles relocated from station P_i to P_j between periods t and $t+\delta_{ij}^t$
Z_i	size of station P_i
a_{i_t}	number of available vehicles at station P_i at the start of period t
$S_{i_t,i_{t+1}}$	number of vehicles stocked at station P_i between periods t and $t+1$
$D_{i_t,j_{t+\delta_{ij}^t}}$	number of user trips from station P_i to station P_j between periods t and $t+\delta_{ij}^t$
P	car-sharing fee per minute driven
C_{mv}	cost of maintenance per vehicle per minute
δ_{ij}^t	travel time between stations P_i and P_j in period t
C_{mp}	cost of maintaining one parking space per day
C_v	cost of depreciation per vehicle per day
C_r	cost of relocation and maintenance per vehicle per period driven

[objective function]

$$\pi = (P - C_{mv}) \sum_{i_t,j_{t+\delta_{ij}^t} \in A} D_{i_t,j_{t+\delta_{ij}^t}} - C_{mp} \sum_{i \in N} Z_i - C_v \sum_{i \in N} a_{i_t} - C_r \sum_{i_t,j_{t+\delta_{ij}^t} \in A} R_{i_t,j_{t+\delta_{ij}^t}}. \tag{1}$$

Equation (1) represents the total profit, which is calculated by subtracting the cost from the revenue. In this study, we aim to maximize the total profit.

In Jorge et al.'s model,

$$a_{i_t} = a_{i_{t-1}} - \sum_{j \in N} D_{i_{t-1},j_{t-1+\delta_{ij}^t}} - \sum_{j \in N} R_{i_{t-1},j_{t-1+\delta_{ij}^t}} + \sum_{j \in N} D_{i_{t-2\delta_{ij}^t},i_t} + \sum_{j \in N} R_{i_{t-2\delta_{ij}^t},i_t}, \tag{2}$$

where a_{i_t} denotes the number of available vehicles at station i at the start of period t.

However, the number of available vehicles was not necessarily the same as the total number of vehicles because the EVs relocated by the staff had to be recharged after the relocation. In this study, we added the following constraint equation:

$$b_{i_t} - \sum_{j_t \in X} D_{i_t,j_{t+\delta_{ij}^t}} - \sum_{j_t \in X} R_{i_t,j_{t+\delta_{ij}^t}} - S_{i_t,i_{t+1}} = 0, \tag{3}$$

where b_{i_t} denotes the total number of vehicles at station i at the start of period t.

3 Numerical Experiments

Numerical experiments were conducted to investigate the effect of the charging time on the revenue. The parameters related to the model settings are listed in Table 2.

The number of customers moving from parking stations i to j was generated using a uniform random number U [0,5]. The cost of car sharing was determined by referring to the prices of Japanese companies that provided car-sharing services [6]. Other cost and vehicle related parameters are listed in Tables 3 and 4, respectively. The parking stations were located in Chitose, Sapporo, and Otaru in Hokkaido, which are well-known tourist destinations (see Fig. 3).

Figure 4 depicts the effect of relocation on charge rate. The horizontal axis in Fig. 4 represents the travel time for relocation, and the vertical axis represents the number of relocations. In this study, we assumed that relocations decrease the charging times of EVs. For example, because the travel time between Sapporo and Otaru is approximately 45 min, we assumed that the charging time of the EV would decrease by 45 min when relocation occurred. Consequently, the recharging rate of the EVs relocated between Sapporo and Otaru became approximately 90%; similarly, the recharging rate of the EVs repositioned relocated Otaru and Chitose was 79%. As the results indicated that the risk of running out of charge increased with repeated relocation of the same EV, it became necessary to further consider the selection and rotation of EVs used for relocation.

Finally, we compared the total profits for three scenarios, namely one without relocation, one with over 45 min of charging time loss (using the numerical example of this study), and one with 15 min of charging time loss. Table 5 presents the average total profits for the three scenarios. The total profit was higher in the scenarios with relocation than that in the one without it. If more than 45 min of relocation time was

Table 2 Parameters

Number of park stations	3
Time frame (min/period)	15
Parking station capacity	20
Initial number of vehicles	10
Number of staffs	3

Table 3 Cost parameters

Fuel cost (yen/min・vehicle)	10
Parking fee (yen/vehicle)	200
Staff fee (yen/min・vehicle)	200

Table 4 Vehicle parameters

Battery recharging rate (%/min)	0.21
Battery consumption rate (%/min)	0.44
Speed (km/h)	45

Fig. 3 Location of the parking station

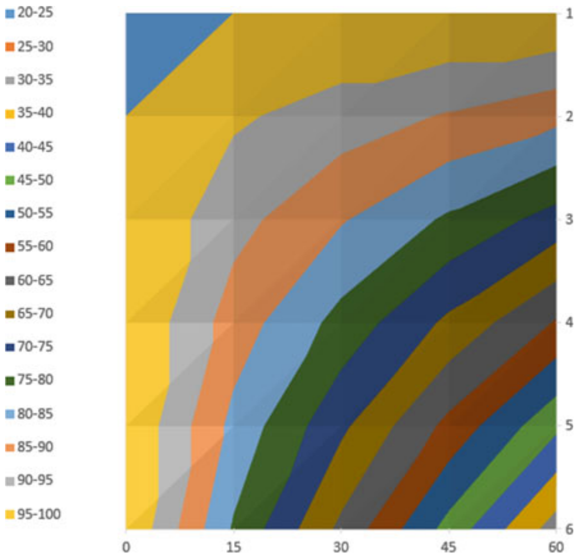


Fig. 4 Recharge rates with relocation time and number of relocations

Table 5 Total profit in the three scenarios

Scenario	Profits
Without relocation	23,490
Relocation (more than 45 min loss)	23,950
Relocation (15 min loss)	25,350

required, the fixed costs associated with the vehicle increased, owing to the effect of the charging time. A 15 min reduction in charging time had no impact on the schedule and increased profits. In this study, cost parameters and recharging time were determined based on the case of only a single company; further research and analysis are needed to determine the impacts of cost and recharging time in greater detail.

4 Conclusions

In this study, we determined the recharging schedule and staff relocation strategy for maximizing profits in a car-sharing system for EVs. The model developed in this study helped to efficiently plan the relocation of staff and EVs, considering the recharging time of EVs.

The experimental results indicated that the relocations decreased the fixed costs of the staff and vehicles, thereby increasing the total profit. If additional charging time was required for relocation and there was insufficient time, the frequency of vehicle relocation needed to be considered to avoid running out of charge. Furthermore, the selection of EVs to be relocated had a significant impact on staff schedules. In our future work, we intend to solve the EV selection problem simultaneously with EV relocation and staff scheduling.

References

1. Knobloch, F., Hanssen, S, V., Lam, A., Pollitt, H., Salas, P., Chewpreecha, U., Huijbregts, M. A. J., Mercure, J. F.: Net emission reductions from electric cars and heat pumps in 59 world regions over time. Nat. Sustain. **3**, 437–447 (2020). https://www.nature.com/articles/s41893-020-0488-7. Accessed 2021 Sept 23
2. car2go Group: SHARE NOW. https://www.share-now.com/. Accessed 2021 Sept 23
3. Renault Group: ZITY. https://en.media.renaultgroup.com/news/zity-groupe-renaults-electric-car-sharing-service-sets-up-in-boulogne-billancourt-1971-989c5.html. Accessed 2021 Sept 23
4. Jorge, D., Correia, G.H.A., Barnhart, C.: Comparing optimal relocation operations with simulated relocation policies in one-way carsharing systems. IEEE Trans. Intell. Transp. Syst. **15**(4), 1667–1675 (2014)
5. Nissan: Nissan LEAF. https://www.nissanusa.com/vehicles/electric-cars/leaf.html. Accessed 2021 Sept 23
6. Nissan e-shair mobi.https://e-sharemobi.com/. (In Japanese). Accessed 2021 Sept 23

Prediction of COVID-19 Hospital Beds by On-Demand Cumulative-Control Analysis

Tetsuo Yamada, Taiga Onodera, Masayuki Matsui, Daisuke Kobayashi, and Eiko Kobayashi

Abstract The COVID-19 virus has been rapidly spreading around the world, and the situation in clinical treatment sites is dynamically changing. In such a situation, the prediction and management of the required number of hospital beds have become issues in order to prevent the collapse of medical care due to the number of patients exceeding the capacity. On the other hand, cumulative flow analysis and on-demand flow management method, which is an extended management system, have been developed and used as a tool for visually managing the state of inventory in factories and warehouses. Onodera et al. (2020) adopted the on-demand cumulative-control method to the estimation of the COVID-19 patient population in a case study. However, the base number of vacant hospital beds is not predicted. This study models the number of COVID-19 hospital beds by using the on-demand cumulative-control method, and estimates required number of vacant beds from the viewpoint of inventory management. Firstly, the cumulative-control analysis is modeled by the input number of hospital beds as the supply, the output number of patients testing positive as the demand, and the number of vacant hospital beds as the number of flow which means the difference between cumulative input and output number. Next, using the on-demand cumulative-control method, the base number of vacant hospital beds is estimated to prevent the overwhelmed healthcare. Finally, a numerical experiment is conducted, and potentials for the appropriate and current management/policy are discussed based on the comparison between the required number of input beds estimated by the on-demand cumulative control method and the actual number of beds.

T. Yamada (✉) · T. Onodera
The University of Electro-Communications, Chofu, Japan
e-mail: tyamada@uec.ac.jp

M. Matsui
Kanagawa University, Yokohama, Japan

D. Kobayashi
Toho University, Ota City, Japan

E. Kobayashi
Japanese Red Cross Medical Center, Tokyo, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_8

Keywords Novel coronavirus · Dynamic inventory control · Bed management · Cumulative flow curve · Case study

1 Introduction

Currently, COVID-19 infection is spreading rapidly in various countries around the world [1], and studies such as simulation and clarification of the virus itself are being actively conducted because of the threat [2]. On the other hand, in clinical treatment sites, there are concerns about medical care collapse due to the ever-changing patient count, the shortage of quarantine facilities, and the prediction and management of the required number of hospital beds [3–7]. However, there has currently been no attempt to use data in real time. Cumulative analysis [8] is a method for visually managing and improving inventory and progress by displaying cumulative inflow and outflow in factories and warehouses on a diagram form. One of the merits by this analysis method is that it enables us to use real time data.

Matsui et al. [9, 10] developed this cumulative analysis method, calculated the reference flow rate using the news boy problem, and developed an algorithm to determine the next supply quantity on demand. Onodera et al. [11] applied the on-demand cumulative-analysis to COVID-19 patient population, where the cumulative number of persons testing positive was predicted by handling the output as the demand. One of the reasons and advantages why the on-demand cumulative-analysis is adapted is to easily forecast COVID-19 patient population by coordinating only two parameters such as exponential smoothing coefficient and period penalty factor. In other words, the number of persons completing treatment was determined and used to predict the cumulative number of persons testing positive. However, it was not used as a supply and demand control method for resources of medical institutions such as hospital beds, medical workers, and pharmaceuticals [12], which were original applications. In addition, rather than using the number of persons completing treatment as the demand, it is better to consider the number of testing positive as the COVID-19 medical demand. Thus, an immediate evaluation may become possible by the clinical site with the medical institution as management.

This study models the number of COVID-19 hospital beds by using the on-demand cumulative-control method [9, 10], and estimates required number of vacant beds from the viewpoint of inventory management. Firstly, the cumulative-control analysis is modeled by the input number of hospital beds as the supply, the output number of patients testing positive as the demand, and the number of vacant hospital beds as the number of flow which means the difference between cumulative input and output number. Next, using the on-demand cumulative-control method, the base number of vacant hospital beds is estimated to prevent the overwhelmed healthcare. Finally, a numerical experiment is conducted and discussed.

2 Methods

2.1 Overview of Procedures

The procedures consisting of 4 steps in this study is as follows. In STEP 1, a state transition diagram is created, and the relationship between COVID-19 and the cumulative analysis are identified. In STEP 2, data such as the number of COVID-19 positive individuals who exclude released patients from medical treatment and death are acquired from public information. In STEP 3, the flow number and lead time of each day are calculated through cumulative analysis, and the problems are extracted. Finally, in STEP 4, we verify whether appropriate management is possible by the on-demand flow rate management method. The following notions are used in this study.

I_t Number of individuals testing positive in period t
$I_t^{'}$ Number of supplied hospital beds in period t (Input supply quantity)
O_t Number of persons who completed their medical treatment in period t
$O_t^{'}$ Number of persons testing positive in period t (Output demand quantity)
X_t Predicted demand in period t
L_t Number of patients (hospital beds used) in period t
$L_t^{'}$ Number of vacant hospital beds in period t (Flow rate)
A_t Total number of hospital beds in period t
N_t Base number of patients in period t (number of hospital beds used)
$N_t^{'}$ Base number of vacant hospital beds in period t (Base flow rate)
α Coefficient of exponential smoothing ($0 \leq \alpha \leq 1$)
$\overline{\beta}_t$ Period penalty factor in period t ($0 \leq \overline{\beta}_t \leq 1$)

2.2 Preparation of a State Transition Diagram Cumulative Analysis

Figure 1 illustrates a state transition diagram for the patients handled in this study. The process from COVID-19 infection to medical treatment first shifts to the test state when the symptoms appear in the infected and non-infected states, and then transitions to a positive or negative state according to the test results. In the case of the positive state, it transitions to the state of cancelled recuperation or death through either hospitalization, recuperation at an institution, or recuperation at home. In the case of the negative state, it becomes the uninfected state, and it transitions to the test state again when the symptom appears.

The cumulative analysis [8] in STEP 3 is a method for grasping the work in process products in factories originally, and the inventory as well as production, sales, and procurement lead times in warehouses and retail stores. Flow rate refers to the work in process, products, or inventory.

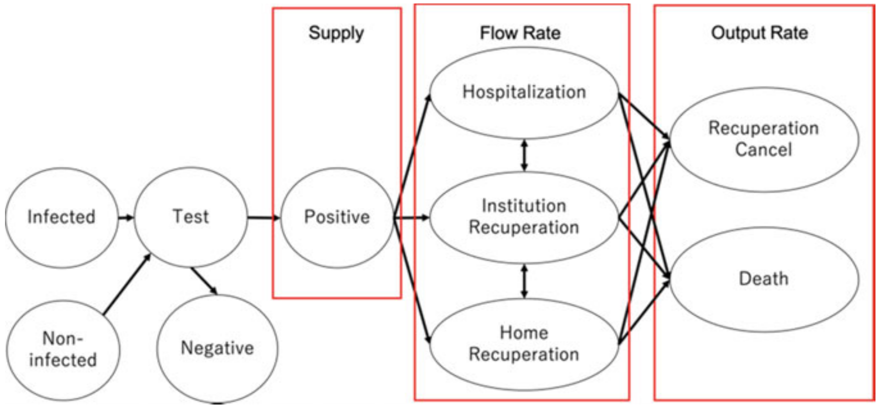


Fig. 1 Transition diagram for the state of infected persons

In the cumulative analysis, a graph of cumulative supply and demand is constructed, with the horizontal axis as the period and the vertical axis as the quantity, as shown in Fig. 3. For instance, the difference in the vertical direction represents the flow rate, and the difference in the horizontal direction represents the lead time. By grasping the flow rate and lead time of each stage, management and improvement of processes and progress can be expected.

The cumulative analysis is performed by handling hospitals and lodging facilities as warehouses. In this study, the on-demand cumulative control method [9, 9] is used as the estimation for the required number of beds in medical institutions rather than the projection of the number of persons testing positive, considering the number of persons testing positive as the demand.

In here, the on-demand cumulative control method is executed by treating the input number of hospital beds as the supply, the number of patients testing positive as the demand, and the number of vacant beds as the flow rate.

2.3 On-demand Cumulative Control Method

The on-demand cumulative control method in STEP 4 used at each process for supply chain management under the on-demand environment developed by Matsui et al. [9, 10]. *On demand* means the delivery of goods and services when requested by customers.

Figure 2 shows a flowchart of the on-demand cumulative control method. In the cumulative control, it is necessary to establish the reference amount of the flow rate for management, that is, the reference flow rate. In the on-demand cumulative control method, this reference flow rate is requested for each period based on the news boy problem so as to minimize the total penalty of the cumulative holding and opportunity loss. This is known as the moving-based inventory.

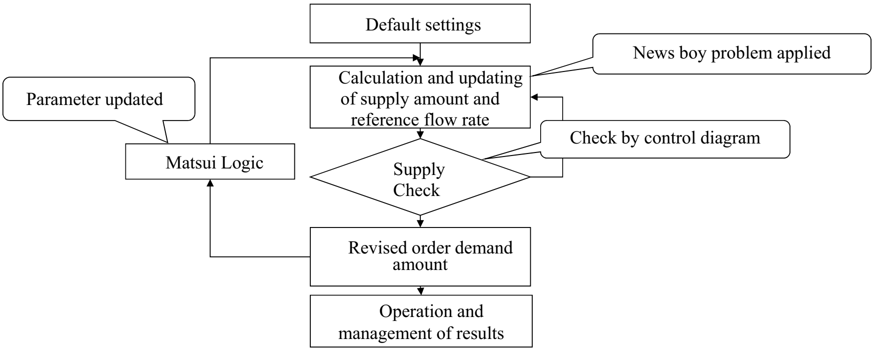


Fig. 2 Flowchart of the on-demand cumulative control method [9]

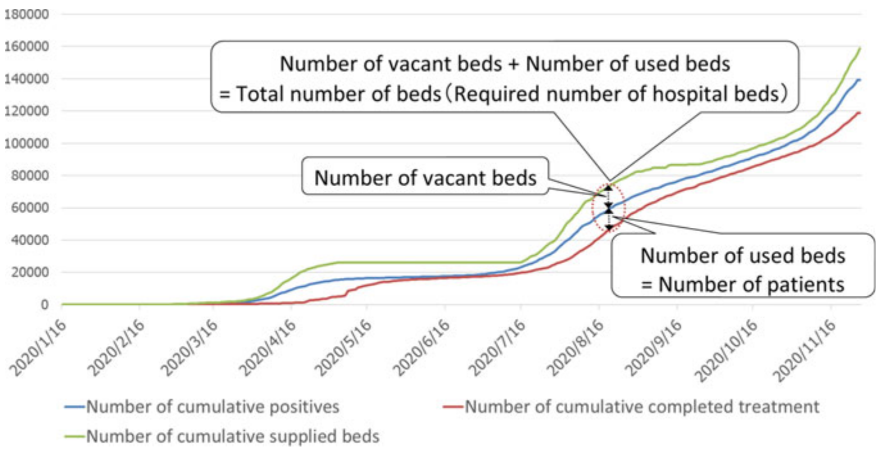


Fig. 3 Example for the number of input hospital beds to COVID-19 patients increased

After obtaining the moving-based inventory, the moving-based inventory I_{t+1} is determined in order that quantity for the supply quantity in the next period L_t would bring flow rate N_t closer to the moving-based inventory I_{t+1}. The indicated supply quantity inflow is determined by taking the demand quantity of the next period into account O_{t+1}.

2.4 Estimation of the Number of Hospital Beds

At this time, Eq. (1) means that for the total number of hospital beds in the period t, A_t, holds the sum of the number of hospital beds used L_t and the number of vacant hospital beds $L_t^{'}$. Here, the number of hospital beds used L_t in Eq. (2) is satisfied with respect to the cumulative number of persons testing positive $O_t^{'}$ and the cumulative number of patients completing treatment O_t because it is the same as the number of patients completing treatment up to the previous study [11]. However, $\sum X_t$ is expressed as the cumulative number of t periods in a certain time series X_t. Similarly, the number of vacant hospital beds $L_t^{'}$ satisfies Eq. (3) for the cumulative number of input beds $I_t^{'}$ and the cumulative number of persons testing positive $O_t^{'}$.

$$L_t + L_t^{'} = A_t \tag{1}$$

$$L_t = \sum O_t^{'} - \sum O_t \tag{2}$$

$$L_t^{'} = \sum I_t^{'} - \sum O_t^{'} \tag{3}$$

From this, when the total number of hospital beds A_t can be expressed by the cumulative number as shown in Eq. (4). Therefore, it becomes possible to determine the required number of hospital beds A_t by determining the number of input hospital beds $I_{t+1}^{'}$.

$$A_t = L_t' + L_t$$

$$= \left(\sum I_t^{'} - \sum O_t^{'}\right) + \left(\sum O_t^{'} - \sum O_t\right) = \sum I_t^{'} - \sum O_t \tag{4}$$

When out-of-stock (opportunity loss) is generated with respect to the number of persons testing positive $O_{t+1}^{'}$ in the number of input hospital beds $I_{t+1}^{'}$, the number of vacant hospital beds $L_t^{'}$ becomes negative. This is considered to be the state of medical collapse. While medical collapse must be prevented, simply increasing the number of vacant hospital beds. However, it may reduce the number of hospital beds for the other than patients instead of COVID-19 ones and maintain vacant hospital beds, which decreases hospital profits. In other words, by solving the news boy problem on opportunity loss (medical collapse) and stock (vacant bed) holding, it becomes possible to manage hospital beds by determining the number of base vacant beds. The cumulative control formula at this time is obtained as Eq. (5). It is noted that $N_t^{'}$ as base number of vacant hospital beds.

$$I_{t+1}^{'} = O_{t+1}^{'} + N_t^{'} - L_t^{'} \tag{5}$$

3 Results

3.1 Example of the Hospital Bed Input

The number of base vacant beds is calculated using the following Eqs. (6)–(8), when the unused bed rate $p = 0.5$ is assumed as constant.

$$N_t^{'} = pA_t = A_t - N_t \tag{6}$$

$$A_t = \frac{1}{1-p}N_t = \frac{1}{1-0.5}N_t = 2N_t \tag{7}$$

$$N_t^{'} = pA_t = \frac{p}{1-p}N_t = \frac{0.5}{0.5}N_t = N_t \tag{8}$$

In this experiment, the on-demand cumulative control method is executed using the cumulative control formula in Eq. (5) and the reference number of vacant beds in Eq. (8). As an example of parameters $\left(\alpha, \overline{\beta}_0\right)$ is set as (0.34, 0.52), where $\overline{\beta}_0$ means an initial period penalty factor $\overline{\beta}_t$ at period 0.

3.2 Example and Discussion for Resulted Cumulative Curve

Figure 3 shows a calculation example for the number of input hospital beds to COVID-19 patients increased using [13]. The difference between the cumulative number of positive persons and that of input hospital beds in Fig. 3 is expressed as the number of vacant beds. On the other hand, while the sum of the number of patients and vacant beds means the required number of hospital beds. In the figure, it can be seen that the number of vacant hospital beds calculated on around July 16 in 2020 is small, but since it is not 0, the medical collapse in the sense of this study has not occurred. Therefore, using the on-demand cumulative control method, it is possible to calculate the number of beds suitable for demand, and more effective bed management is expected rather than simply preparing the number of beds that will be two times higher than the number of current patients (the idle (unused bed) rate is 50%).

As described above, it seems to be possible to support a decision-making whether the current management and policy are appropriate by comparing the number of input beds estimated via the on-demand cumulative control method with the actual number of beds. For example, if the current number of beds is less than the estimated number of beds, it is possible to encourage improvement in treatment such that the actual number of beds and the number of employees are increased by subsidy to each facility. On the other hand, if it is understood that the present value is large, it becomes possible to establish policies such as maintaining the status quo.

However, the number of hospital beds in this result means the number in the entire isolation facility, and the number of hospitals, accommodation facilities, homes, and so on, have not been taken into account. Therefore, if the number of isolation facilities and hospital beds can be calculated in the future, it will be possible to use these values as indices of more accurate management, and that will be a future issue.

4 Summary and Future Studies

This study modeled the number of COVID-19 hospital beds by using the on-demand cumulative-control method [9, 10], and estimated the required number of vacant beds from the viewpoint of inventory management. Firstly, the cumulative-control analysis was modeled by the input number of hospital beds as the supply, the output number of patients testing positive as the demand, and the number of vacant hospital beds as the number of flow which meant the difference between cumulative input and output number. Next, using the on-demand cumulative-control method, the base number of vacant hospital beds was estimated to prevent the overwhelmed healthcare. Finally, a numerical experiment was conducted and discussed. It demonstrated that the on-demand cumulative control method enabled us to support a decision-making whether the current management and policy were appropriate or not.

Future studies should verify whether the extracted problem can be dealt with by the on-demand cumulative control method, continue to investigate the present state of data and clinical treatment sites, compare to the other models which forecast patient populations, etc.

Acknowledgements This research was partially supported by the Japan Society for the Promotion Science (JSPS), KAKENHI, Grant-in-Aid for Scientific Research (C), JP19K04915 from 2019 to 2023 and Grant-in-Aid for Scientific Research (A), JP18H03824, from 2018 to 2023.

References

1. Mizutani, T.: COVID-19-accurate knowledge to control threats. Tokyo Kagaku Dojin (2020). (In Japanese)
2. Kurahashi, S.: Estimating effectiveness of preventing measures for 2019 Novel Coronavirus diseases (COVID-19). In: 9th IIAI International Congress on Advanced Applied Informatics (IIAI AAI 2020), pp. 491–496 (2020)
3. Tsuchiya, T.: A study of spreading COVID-19 infections. GRIPS Policy Reserch Center, 20–04 (2020). https://doi.org/10.24545/00001758. (In Japanese)
4. Muto, M.: Medical sytstem will change by COVID-19. Japan Medical Publisher, 18–31 (2020). (In Japanese)
5. Specified non-profit corporation for Local Medical Care / Nursing Care Study Group JAPAN, Healthcare System Research Institute Co., Ltd.: Fight against COVID-19, 120-days record of on-site doctors. PHP Editors Group (2020). (In Japanese)

6. Nikkei Inc.: Insufficient beds, fear of up to 65,000 beds; Anxiety about preparing for the second wave of infection https://www.nikkei.com/article/DGXMZO60947120Z20C20A6EA2000/. Accessed on Sep. 22, 2020. (In Japanese)
7. GHC Evolving Media for Medicine: The third wave of the COVID-19, if the number of infectious diseases is not reduced, regional medical care will collapse, and the medical plan describes "securing spare beds, education for staff", etc. https://gemmed.ghc-j.com/?p=37582. Accessed on 5 Jan 2021. (In Japanese)
8. Usuki, J., Kitaoka, M.: Classification of cumulative curve from initial inventory and relationship between inventory quantity and lead time. Jpn. Ind. Manag. Assoc. **51**(6), 558–565 (2001). https://doi.org/10.11221/jima.51.558. (In Japanese)
9. Matsui, M., Uchiyama, H., Hujikawa, H.: Progressive-curve-based control of inventory fluctuation under on-demand SCM. Jpn. Ind. Manag. Assoc. **56**(2), 139–145 (2005). https://doi.org/10.11221/jima.56.139. (In Japanese)
10. Matsui, M., Hujikawa, H., Ishii, N.: Supply-to-Demand Management -Towards Post ERP/SCM. Asakura Publishing Co., Ltd. (2009). (In Japanese)
11. Onodera, T., Yamada, T., Matsui, M., Kobayashi, D., Kobayashi, E.: An application of on-demand cumulative-control analysis to COVID-19 patient population. J. Soc. Plant Eng. Jpn. **33**(4), 1–9 (2022). (In Japanese)
12. Ministry of Health, Labor and Welfare: Overview of the (dynamic) survey of medical institutions and hospital report (2019). https://www.mhlw.go.jp/toukei/saikin/hw/iryosd/19/dl/03byouin01.pdf. Accessed on 5 Jan 2021. (In Japanese)
13. Ministry of Health, Labor and Welfare: Novel Coronavirus. https://www.mhlw.go.jp/stf/covid-19/open-data.html. (In Japanese). Accessed on 28 Nov 2020

Digital and Cyber Manufacturing and Services

A Study on Skill Transfer Using Augmented Reality

Koichi Akagi, Noboru Hayakawa, Takeshi Morishita, and Manabu Sawaguchi

Abstract Effective methods of skill transfer are required at the manufacturing site of heavy industry with high-mix, low-volume manufacturing. Several previous studies have shown that giving unskilled worker simultaneous experiences in the skilled worker's visual sense from a first-person perspective and force sense is effective for the skill transfer. This paper proposes a method of skill transfer by creating holographic moving images of workpieces and tools from the skilled worker's motion. In order to analyze the effectiveness of this method, a control experiment by the assembly work of Tetris Block was carried out as a preliminary experiment. By using the proposed method and movie instruction which has already been put to practical use, each experimental participant's assembly work was observed and compared, and the difference of two methods was analyzed from the three viewpoints of the work procedure, work time, and body motion. As a result, the effectiveness of the proposed method was confirmed from three perspectives; the margins of the body motion can be provided by the movement of holograms, the movement of holograms can communicate the skilled worker's detailed motion information, and the skill transfer is possible in the virtual space environment which is not affected by real space. The future task is to verify the findings of this study by applying proposed method to actual field works.

Keywords Skill transfer · Augmented reality · Control experiment

---

K. Akagi
Research Organization of Open Innovation and Collaboration, Ritsumeikan University, Osaka, Japan
e-mail: koichi_akagi@icloud.com

N. Hayakawa · T. Morishita (✉)
Mitsubishi Heavy Industries, Ltd., Hyogo, Japan

M. Sawaguchi
Graduate School of Technology Management, Ritsumeikan University, Osaka, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_9

1 Introduction

In the manufacturing field of heavy industry which is high-mix, low-volume manufacturing, the skills highly dependent on the individual ability and not easy to transfer to unskilled workers by the instruction with formal language and diagrams are required. The purpose of this study is to verify the effectiveness of utilizing augmented reality to effectively transfer those skills. To this end, the items necessary for the skill transfer are clarified through previous researches, and a method of using augmented reality that satisfies those items is proposed. Then, the proposed method is analyzed by a control experiment, and finally the effectiveness is examined.

2 Previous Study

Polanyi mentions that the skill transfer is getting a sense of skill by repeating exploratory internalization of the "marginal control for the motion of skills" in the skilled workers' and unskilled workers' brain [1]. On the other hand, the following attempts to utilize ICT for the skill transfer are discussed. In order to confirm the effect of first-person vision on training through the training of the cardiovascular resuscitation method, Yonemura et al. conducted training to imitate the motion of the trainer with the same viewpoint as the trainer by showing a recorded trainer's view image (first-person vision) in the view image of the experimental participants with video pass-through HMD (Head Mount Display). Based on the results, it is argued that the sharing of the first-person perspective is effective for training [2]. Watanuki and Kojima discussed that the transfer of skills in casting mold assembly is possible by providing a place where workers can experience the visual sense and force sense (a sense of human feeling when contacting the object) with VR technology and dynamic presentation equipment Hiyama et al. [3]. Hirose have constructed system with wearable computers that allows to experience and learn the visual, hearing, and force senses of the skilled workers, and conducted evaluation experiments in some processes of papermaking, which is the traditional skills. As a result, it is confirmed that the best way to imitate the work of the skilled worker is learning by the simultaneous experience in the visual sense from a first-person perspective and force sense [4]. Thus, it has become clear that the unskilled worker can learn the "marginal control for the motion of skills" mentioned by Polanyi in previous studies by the visual sense from a first-person perspective and actual force sense. However, there are no research paper mentioning the skill transfer method using newly developed augmented reality technology.

Fig. 1 Hologram displayed by augmented reality headset

3 Proposed Method and Hypothesis

As a method to utilize the augmented reality in order to provide unskilled worker with the skilled worker's visual sense from a first-person perspective and force sense, the following method is proposed. The hypothesis that the skills can be transferred more effectively than the conventional method by applying this method.

Please refer to Fig. 1. HMD-type display device (Microsoft HoloLens) is used as an augmented reality display device.

Unskilled workers look at the holograms of 3D model of workpieces and tools displayed in HMD from a first-person perspective. These holograms simulate the movement of the workpieces and tools which the skilled worker is moving during the work. Unskilled workers experience the force sense that skilled workers feel in the real work by moving actual workpieces and tools in the same manner as the holograms displayed in the HMD.

4 Analysis Method

As a control experiment, the assembly work of the Tetris Block shown in Fig. 1 is carried out. Six graduate students without experience in assembly of Tetris Block are selected as experimental participants ("a" to "f"). These participants carry out the assembly work of Tetris Block three times in a row with the following instruction method; participants "a" to "c" with the movie instruction method which haves already been put into practical use as skill transfer method, and other participants "d" to "f" with the proposed augmented reality method. As shown in the photograph in Fig. 2, work of these participants is recorded and analyzed with two cameras for movie recording.

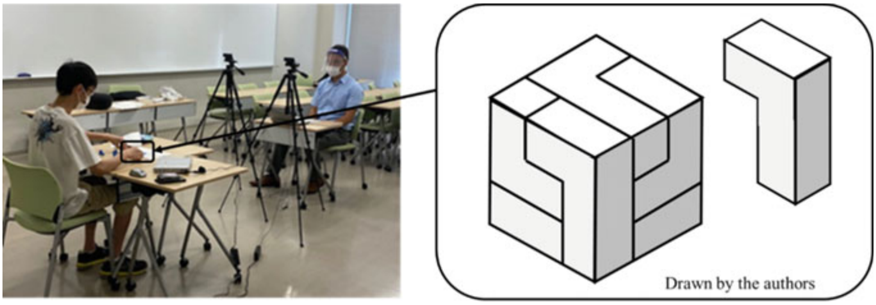


Fig. 2 Tetris block and experimental scene

Fig. 3 Displayed image and participants of movie instruction method on PC (left) and augmented reality (HMD) method (right)

4.1 Movie Instruction Method (See the Left Image in Fig. 3)

The visual image from a first-person perspective of the assembly work is displayed on the screen of the PC on the desk. The skilled worker's hand and blocks are shown in the movie. The work instructions are displayed as a subtitle at the bottom of the screen and the audio instructions are played.

4.2 Augmented Reality Instruction Method (See the Right Image in Fig. 3)

The movie and holograms are shown to participants with HMD. The same movie as movie instruction method is displayed ahead of the virtual space.

On the other hand, the holograms of the block from a first-person perspective are repeatedly shown the same motion as the skilled worker's motion (once every 6 s).

Table 1 Observation results of work procedure

Participants	Movie instruction			Augmented reality		
	a	**b**	**c**	**d**	**e**	**f**
1st assembly	**N**	**N**	**N**	**N**	**M**	**N**
2nd assembly	**N**	**C2**	**C1**	**N**	**N**	**N**
3rd assembly	**N**	**C3**	**N**	**N**	**N**	**N**

N Followed the standard work procedure
C1 Changed on purpose (performed processes 1 and 2 at the same time)
C2 Changed on purpose (performed processes 6 and 7 at the same time)
C3 Changed on purpose (combination of C1 and C2)
M Incidentally changed

5 Analysis Results

The results of analysis on the difference between two instruction methods based on the experiment videos from three viewpoints of work procedure, work time, and body motion.

5.1 Analysis of Work Procedure

The observation was carried out to check whether the participants follow the work procedure (total of 7 processes) and the results were categorized and summarized in Table 1. With the movie instruction method, there were some cases where the work procedure was intentionally changed for the purpose of improving work efficiency in participant's decision. Incidentally, participants were not asked to follow the work procedure.

5.2 Analysis of Work Time

Table 2 shows the results of an analysis of the time it took for the participant from the start of motion to completion of assembly for each process during the third assembly. Each process described in Table 2 has almost same movement of taking block from the work desk and assemble. The time for the work that has changed the work procedure on purpose (participant "b") is not adopted. It also shows the average and standard deviation of the total work time. As a result, participants instructed by augmented reality have shorter average work time and smaller standard deviation.

Table 2 Comparison of work time (3rd assembly) unit: second

	Movie Instruction			Augmented reality		
Participants	A	b	c	d	e	F
Place block 1	9.43	N/A	0.73	3.46	1.31	2.54
Assemble block 2	14.34	5.04	55.46	3.16	2.23	3.19
Assemble block 3	7.38	5.98	7.71	5.33	2.89	4.98
Assemble block 4	10.96	5.61	12.16	7.11	5.24	5.84
Assemble block 5	4.69	N/A	49.03	7.76	5.71	4.01
Assemble block 6	2.51	N/A	16.39	4.83	2.36	4.01
Average work time	8.22	N/A	23.58	5.28	3.29	4.10
Standard deviation	4.29	N/A	22.89	1.87	1.77	1.19

5.3 Analysis of Body Motion

The skeleton position (in a coordination system composed of two camera viewpoints) of each participant is estimated using VISIONPOSE (a commercial software that can estimate the human skeleton from the movie) from the recorded work videos. These skeleton positions are converted into a coordinate system from a first-person perspective (participant-centered coordinate system) as shown in Fig. 4. The origin of the coordinate system is set to the position of chest spine, the direction of X' axis is set to the direction of both shoulders.

Next, the trajectory of each skeleton is calculated, and the parts which clearly expresses the participant's motion are confirmed from those results. Figure 5 shows the trajectory of both elbows of participant "d" (see Fig. 4) during Process 4 "Assemble block 3" at the third assembly with the augmented reality instruction method and it clearly expresses the motion of the participant.

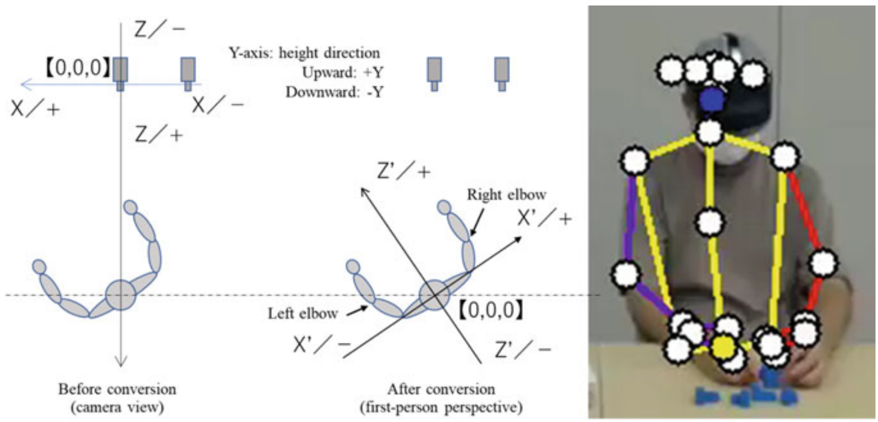


Fig. 4 Coordinate conversion and distance of both elbows

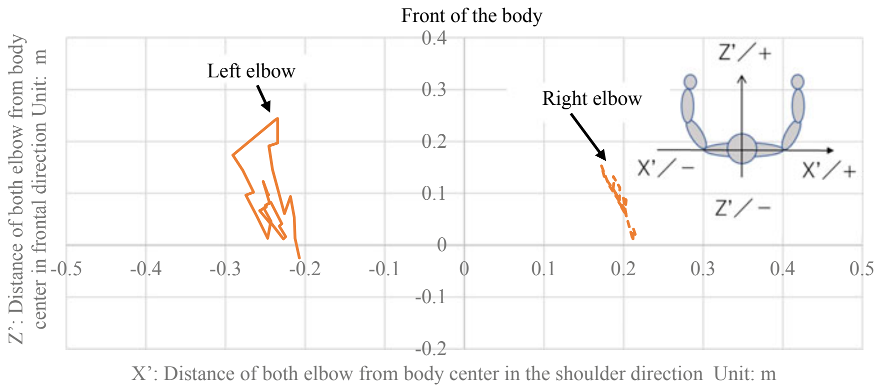


Fig. 5 Position of both elbows (coordinate system from a first-person perspective)

With the trajectories of both elbows, the body motion of each participant is compared. In particular, the distance from the center of participant's body to both elbows is calculated from the start of motion to completion of assembly during the process of "Assemble block 3" at the third assembly (see Fig. 6a, b). There is a variation in posture among the participants with the movie instruction method (there is a large difference between the distance from the body center to both elbows, and this difference varies depending on the participants). Also, there is a variation in the clarity of the motion of the left elbow when picking up the block (participant "b" does not show the clear motion of the left-hand). On the other hand, with the augmented reality instruction method, there is no noticeable difference in the posture and the clarity of the motion among the participants.

6 Analysis and Discussion

The effectiveness of the augmented reality instruction method is considered based on the analysis results of three viewpoints of work procedure, work time, and body motion.

Firstly, the analysis result in viewpoint of work procedure is described. Participants instructed by augmented reality followed the work procedure without thinking, while participants instructed by movie changed the work procedure on purpose. The reason for this is that participants instructed by augmented reality are presumed to try to follow the margins of motion given by the movement of holograms from a first-person perspective. This confirms the effectiveness of the hologram to provide the "marginal control for the motion of skills" mentioned by Polanyi, that of to say, the margins of motion.

Secondly, the analysis result in viewpoint of work time is described. Work time of participants instructed by augmented reality is shorter and less variation than

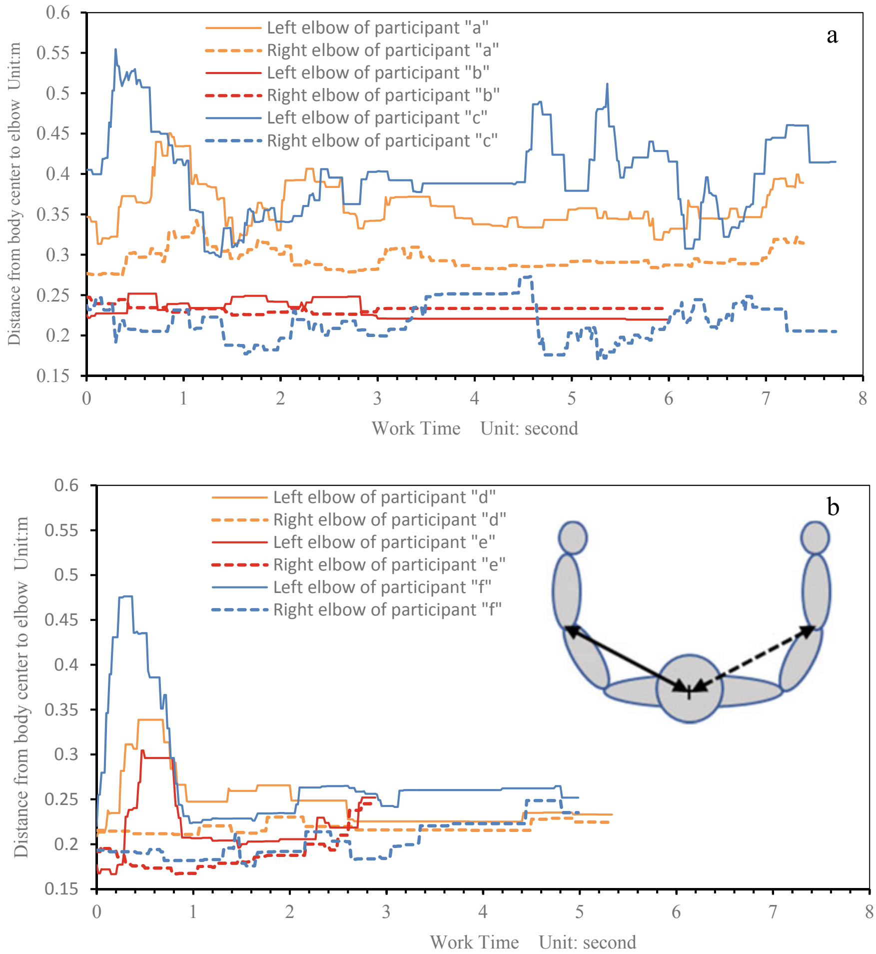


Fig. 6 **a** Distance from body center to both elbows of participants instructed by using movie. **b** Distance from body center to both elbows of participants instructed by using augmented reality

that of participants instructed by movie. The reason for taking long time was that the participant couldn't assemble blocks correctly and struggled. This event also happened to participants "e" and "f", who were instructed by augmented reality, at the first assembly; however, this did not happen afterward. Incidentally, the participant "c", who was instructed by movie, experienced the same event even at the third assembly. This confirms the effectiveness of transferring detailed information on the motion of skilled worker through holograms.

Finally, the analysis result in viewpoint of body motion is described. There is less difference in the posture and the clarity of the motion among the participants instructed by augmented reality than participants instructed by movie. This is because participants instructed by augmented reality take the same posture and motion as a skilled worker since they try to follow the movement of holograms in the virtual space from a first-person perspective. On the other hand, participants instructed by movie are influenced on their posture and motion by the constraints of the position of the computer on the desk in the real space. Therefore, the effectiveness of transferring the motion in the virtual space was conformed.

7 Conclusion

It was suggested that the proposed augmented reality instruction method is effective for the skill transfer from three viewpoints. With the knowledge obtained in this study, we are set to perform further experiment with more participants and apply the proposed method to the skill transfer of the actual manufacturing work and verify the effect. The academic contribution of this study is to show the effectiveness of skill transfer by overlapping the real blocks and the movement of holograms provided by augmented reality.

The social contribution of this study is to provide a concrete method to utilize the augmented reality for the skill transfer. By applying this method to domestic and overseas manufacturing sites, improvement of quality and effective utilization of human resources can be expected.

References

1. Polanyi, M.: The Tacit Dimension. The University of Chicago Press (1966)
2. Yonemura, T., Hashimoto, Y., Kondo, D., Niwa, M., Iizuka, H., Ando, H., Maeda, T.: Effects of CPR training using view sharing system. Trans. Virtual Reality Soc. Jpn. **16**(4), 623–632 (2011). https://doi.org/10.18974/tvrsj.16.4_623
3. Watanuki, K., Kojima, K.: Knowlegde acquisition and job training for advanced technical skills using immersive virtual environment. J. Adv. Mech. Des. Syst. Manuf. **1**(1), 48–57 (2011). https://doi.org/10.1299/jamdsm.1.48
4. Hiyama, A., Doyama, Y., Miyashita, M., Ebuchi, E., Seki, M., Hirose, M.: Artisan skill training system using multisensory expreiance data from first-person standpoint. Trans. Virtual Reality Soc. Jpn. **16**(4), 643–652 (2011). https://doi.org/10.18974/tvrsj.16.4_643

Humanized Robot of New Method and Time System and Its Management: A Digital Transformation Case of Convenience Store Type

Masayuki Matsui, Eri Ohto-Fujita, and Nobuaki Ishii

Abstract As the world of artifacts becomes unstable, today the nature of artifacts and their ideal form are being questioned. The world of artifacts is comprised of human, material, money, and information (3 M&I) systems, and it consists of the balance between companies, households, the economy and its surroundings (the environment). This paper continues from the earlier publications since 2016 and proposes a DX-design method for the realization of a cockpit-type and demand-to-supply H = W robot based on artifact principles. In recent years, we have entered a second era, in which managerial enterprise bodies have come to operate like clockwork and in which resources are moved. Within this precise managerial integration, by looking at the nature of waste versus efficiency in clock systems, a dilemma (Nash's zone) can be seen between knowledge system integration (waste, *muda*) versus sharing (efficiency). On the other hand, the modern industrial engineering (IE) of the method study that forms the crux of this time study is regarded to be mathematical programming (MP). Since the drifting management era, manufacturing has involved capital and labor, but in the newly manifest knowledge society, the means (methods) of production are shifting towards knowledge creation. Modes of management and integration are also made possible mainly through the workers who possess such knowledge. From the above, the humanized managerial enterprise body can be considered as being both cockpit-type and as a demand-to-supply corporate clock through the H = W-type embodiment of a humanized artifact. In this study, we propose a next-generation method and a new method of time study. Moreover, this new method is formulated using Matsui's matrix method of introduction-development-transformation-conclusion-BG-type story. The existence (how it is), and the purpose (what form it should take) of the exemplar artifact, which in this

M. Matsui (✉)
The University of Electro-Communications, Tokyo, Japan
e-mail: matsui@se.uec.ac.jp

E. Ohto-Fujita
University of Agriculture and Technology, Tokyo, Japan

M. Matsui · N. Ishii
Kanagawa University, Yokohama, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_10

case is an enterprise robot, are clarified, forming a basis for future discussions, and which may provide a forum for debate going forward.

Keywords Humanized artifacts · Method and time · Digital transformation (DX) · Matsui's matrix method

1 Introduction

As the world of artifacts becomes unstable, today the nature of artifacts and their ideal form are being questioned. The world of artifacts is comprised of humans, material, money, and information (3M&I) systems, and consists of the balance between companies, households, the economy, and its surroundings (the environment). However, friction can be seen between the two systems, upper and lower, observed in nature versus artifacts science.

From recent research [1, 2] into artifact science and visualization (robotization) beginning from factory physics and science [3], it has been discovered that the pair-map form (invented in 1983 [4]) is a microcosm of nature versus artifacts body, and the duality therein contains three principles, i.e., balancing, sandwiching, fractal.

This paper continues from the earlier project in 2016 and proposes a DX-design method for the realization of a cockpit-type demand-to-supply H = W robot based on artifact principles [5]. To sketch out the concept, the "H = W system" considers a demand-to-supply robot consisting of two systems, a heart (H) and waist (W), where W is a resultant "stable equilibrium," and H is an "active control-valve" in constant operation.

On the basis of the H = W system, we propose and develop the new method and time-typed management beyond Taylor's scientific management [6]. In particular, the humanized robot of a convenience store type would be formulated by Matsui's matrix equation (Matsui ME), and its information flow of PDCA-CAPD versus IDTC-BG logic would be discussed.

2 Development of a Humanized Enterprise Robot

2.1 A New Clock Type Management System

In humanity's early days, time meant the rising and setting of the sun marking day and night. Since the Industrial Revolution, however, "minutes" have come into focus. In recent years, we have entered a second era, in which managerial enterprise bodies have come to operate like clockwork and in which resources are moved. Within this precise managerial integration, by looking at the nature of waste versus efficiency in clock systems, we can see a dilemma between knowledge system integration (waste) versus sharing (efficiency) [7].

On the other hand, the modern industrial engineering (IE) of the method study that forms the crux of this time study is regarded to be mathematical programming (MP). Since the drifting management era, manufacturing (the means of production) has involved capital and labor, but in the newly manifest knowledge society, the means (methods) of production are shifting towards knowledge creation. Modes of management and integration are also made possible mainly through the workers (actors) who possess such knowledge.

From the above, the humanized managerial enterprise body can be considered as being both cockpit-type and demand-to-supply corporate clock-like through the H = W-type embodiment of a humanized artifact. Using a convenience store as an example of such an enterprise, sales for a 24-h period are managed using the long hand of a clock, and if daily inventory is replenished only once at midnight, then decrease in inventory due to demand can be managed using the short hand of a clock.

In the course of this study, we next propose a next-generation method and a method for time study. Moreover, this new method is formulated using Matsui's matrix method of introduction-development-transformation-conclusion-BG-type story. The existence (how it is), and the purpose (what form it should take) of the exemplar artifact, which in this case is an enterprise robot, are clarified, forming a basis for future discussions, and which may provide a forum for debate going forward.

2.2 A Method and Time Management System

For the convenience store, the performance could be traced by the clock model and its wave equation, and it is possible to formulate it by PDCA cycle [8] in Fig. 1. Figure 1 shows an example of *Jidouka* management type for the convenience store robot.

For Fig. 1, the shorthand of the upper level monitors the process of a 24 h (360°) cycle in revenue, and another long hand traces those in inventory at the convenience display. The convenience store robot pursues cyclically the goal of ER (EN) = LT in the pair-map and could maximize the fund of economics (ER) × reliability (LT).

3 Managerial Formulation Based on Matsui's Matrix Method

3.1 Matsui's IDTC-BG Formulization

Generally, the Matsui's matrix method could utilize the input–output representation by the pair-matrix and its chain. This pair-matrix consists of the pair-elements with input–output relationship of objects in the 3 M&I variety, and could be available to develop a fractal (nested) chain story of the IDTC-BG type.

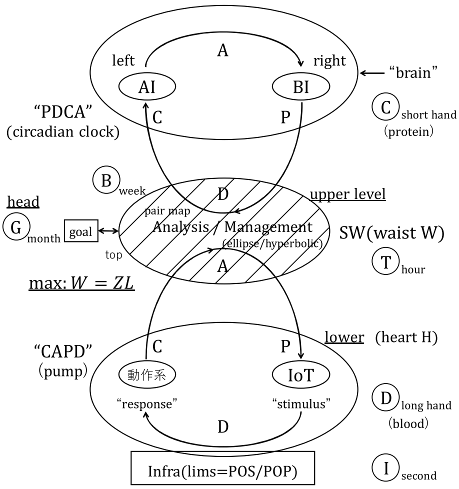


Fig. 1 An outline of a humanized robot and its method of time management system

This type is a chain of I (introduction), D (development), T (transformation), C (conclusion), B (balance), and G (goal), and Matsui's matrix equation (Matsui's ME) is usually represented by matrix notation as follows [3]:

$$\boldsymbol{g} = y^T D^T T^T C^T B^T, \text{ Type } I \tag{1}$$

where y is an input vector, and $\boldsymbol{g}$ is a goal vector.

Now, let us develop the Matsui's matrix equation of a convenience store type by $x = Dy$, $z = Tx$, $u = Cz$ and $v = Bu$ as follows:

Introduction (I) and development (D) stages

At the Eq. (1), the vector x is from the definition: $x = Dy$ as follows:

$$\begin{matrix} & \textbf{< I >} & \textbf{< D >} & \\ \boldsymbol{x^T} = \boldsymbol{y^T D^T} = (d_t \quad d_{t+1} & & \overset{m}{\smile} & \\ & d) & \left|\, --- (d, O) \right. & (2\ ) \end{matrix} \tag{2}$$

For this Eq. (2), the relation of next input (o) and next demand ($d \rightarrow m$) could be represented by using the standard inventory (MI) and end inventory (I). That is, at the development (micro) stage, x = Dy, and is obtained as the following progressive equation from $d_t(d_{t+1})$ to $O_t(O_{t+1})$:

$$\begin{pmatrix} \overline{O}_t \\ \hat{O}_t \end{pmatrix} + \begin{pmatrix} I_{t-2} \\ I_{t-1} \end{pmatrix} = \begin{pmatrix} d_t \\ d_{t+1} \end{pmatrix} + \begin{pmatrix} MI_t \\ MI_{t+1} \end{pmatrix} \tag{3}$$

where d is a forecasting, $m(O)$ is actual (order) demand, and both are usually assumed to be equivalent.

Conclusion (C) and balance (B) stages: At the conclusion stage, there is u = Cz as follows:

$$\boldsymbol{u}^T = \boldsymbol{z}^T \boldsymbol{C}^T = (ERLT)\begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix}. \tag{4}$$

Equation (4) corresponds to the rotation of axis (correlation θ).

Also, at the balance stage, there is v = Bu. If the lead time is decomposed of $LT = m(Lm) + D(LD)$, the following progressive equation from $ER_t(LT_t)$ to $ER_{t+1}(LT_{t+1})$ is obtained:

$$\underset{\text{Kai} \to A'}{\begin{pmatrix} ER_{t+1} \\ LT_{t+1} \end{pmatrix}} + \underset{B}{\begin{pmatrix} EC \\ LD \end{pmatrix}} = \underset{A}{\begin{pmatrix} ER_t \\ LT_t \end{pmatrix}} + \underset{C}{\begin{pmatrix} MP \\ L_m \end{pmatrix}} \tag{5}$$

This progressive Eq. (5) shows the lifecycle of protein Kai A', A, B and C in the Circadian clock [9]. That is, the Eq. (5) is similar to the Eq. (6):

$$\text{Kai}A' = \text{Kai}A + \text{Kai}C - \text{Kai}B,\ \text{protein} \tag{6}$$

Goal (G) stage: Finally, there is a goal vector, $\boldsymbol{g}$ = (ER LT). We would pursue the goal of a convenience store robot by the following equation:

$$\text{fund: } W = ER_{t+1}(Z) \times LT_{t+1}(\overline{L}) \to \max \tag{7}$$

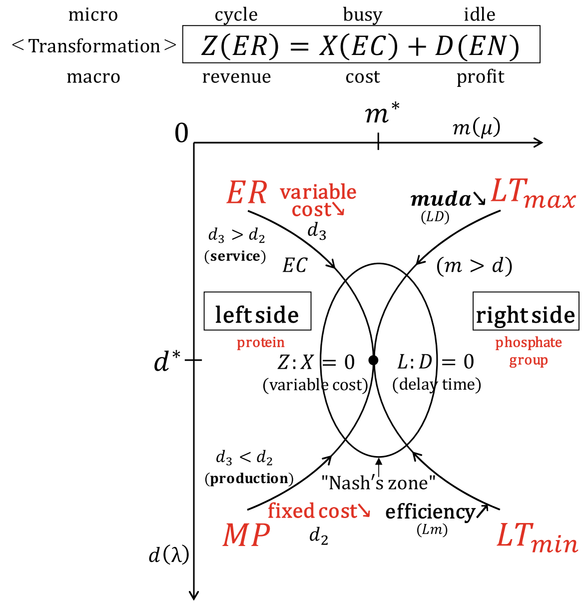
at the next (t + 1) goal, $\boldsymbol{g} = (ER_{t+1}\ LT_{t+1})$.

3.2 Digital Transformation (DX) Issues on Pair Map

At the middle (transformation) stage, there is Z = Tx on the hyperbolic-typed pair map shown in Fig. 2. This shows that the lead time (LT) is convex at the right side, and the revenue (ER) is concave at the left side. The goal of the convenience store robot is located in the neighborhood of the center in the pair map and is pursued at the bipolar type by the one-leaf hyperboloid of nature vs. artifacts body.

For Fig. 2, Fig. 3 could give the evidence example of a job shop case. Figures 2 and 3 show that the pair map microcosms consist of the cross ellipse, pair hyperbolic and maximal profit zone, and this zone would be similar to the Nash's class of so-called bankruptcy, black hall and marginal diversity at the similarity's dilemma of nature vs. artifacts body.

Fig. 2 Outline of hyperbolic-typed pair map with Nash's zone at H = W (waist)



4 Matsui's IDTC-BG and Its Management

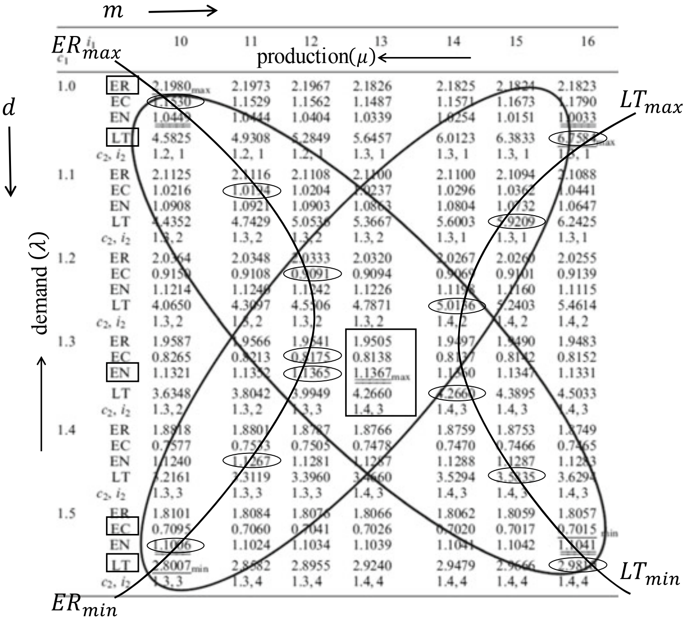
4.1 Objective Formulation of IDTC-BG Type

The Matsui's objective formulation and its management method (tool) of enterprise robot type would now be available. This management utilizes the eigenvalue of pair matrix later. From 3.1 and 3.2, we could totally obtain the objective formulation of IDTC-BG type at a convenience store as follows:

Figure 4 shows the Matsui's matrix equation of IDTC-BG type and its recursive type management at the convenience store. That is, for Fig. 4, the transformation (T) corresponds to the eigenvalue stage $\langle \boldsymbol{e}_1 \boldsymbol{e}_2 \rangle$, and the following contraction is obtained from Fig. 4 and its development [2]:

$$g^T = (\boldsymbol{e}_1 ER_t + \boldsymbol{e}_2 ER_{t+1} \quad \boldsymbol{e}_1 LT_t + \boldsymbol{e}_2 LT_{t+1})^T. \tag{8}$$

Then, Eq. (8) means that the next goal becomes the weighted mean of present and forecasting periods, and thus, the look-ahead control (management) is automatically possible toward the successive goals.



c_1	i_1	10	11	12	13	14	15	16
1.0	ER	2.1980max	2.1973	2.1967	2.1826	2.1825	2.1824	2.1823
	EC	1.1530	1.1529	1.1562	1.1487	1.1571	1.1673	1.1790
	EN	1.0449	1.0444	1.0404	1.0339	1.0254	1.0151	1.0033
	LT	4.5825	4.9308	5.2849	5.6457	6.0123	6.3833	6.7584max
	c_2, i_2	1.2, 1	1.2, 1	1.2, 1	1.3, 1	1.3, 1	1.3, 1	1.3, 1
1.1	ER	2.1125	2.1116	2.1108	2.1100	2.1100	2.1094	2.1088
	EC	1.0216	1.0194	1.0204	1.0237	1.0296	1.0362	1.0441
	EN	1.0908	1.0921	1.0903	1.0863	1.0804	1.0732	1.0647
	LT	4.4352	4.7429	5.0536	5.3667	5.6003	5.9209	6.2425
	c_2, i_2	1.3, 2	1.3, 2	1.3, 2	1.3, 2	1.3, 1	1.3, 1	1.3, 1
1.2	ER	2.0364	2.0348	2.0333	2.0320	2.0267	2.0260	2.0255
	EC	0.9150	0.9108	0.9091	0.9094	0.9069	0.9101	0.9139
	EN	1.1214	1.1240	1.1242	1.1226	1.1198	1.1160	1.1115
	LT	4.0650	4.3097	4.5506	4.7871	5.0166	5.2403	5.4614
	c_2, i_2	1.3, 2	1.3, 2	1.3, 2	1.3, 2	1.4, 2	1.4, 2	1.4, 2
1.3	ER	1.9587	1.9566	1.9541	1.9505	1.9497	1.9490	1.9483
	EC	0.8265	0.8213	0.8175	0.8138	0.8137	0.8142	0.8152
	EN	1.1321	1.1352	1.1365	1.1367max	1.1360	1.1347	1.1331
	LT	3.6348	3.8042	3.9949	4.2660	4.2660	4.3895	4.5033
	c_2, i_2	1.3, 2	1.3, 2	1.3, 3	1.4, 3	1.4, 3	1.4, 3	1.4, 3
1.4	ER	1.8818	1.8801	1.8787	1.8766	1.8759	1.8753	1.8749
	EC	0.7577	0.7533	0.7505	0.7478	0.7470	0.7466	0.7465
	EN	1.1240	1.1267	1.1281	1.1287	1.1288	1.1287	1.1283
	LT	3.2161	3.3119	3.3960	3.4660	3.5294	3.5835	3.6294
	c_2, i_2	1.3, 3	1.3, 3	1.3, 3	1.4, 3	1.4, 3	1.4, 3	1.4, 3
1.5	ER	1.8101	1.8084	1.8076	1.8066	1.8062	1.8059	1.8057
	EC	0.7095	0.7060	0.7041	0.7026	0.7020	0.7017	0.7015min
	EN	1.1006	1.1024	1.1034	1.1039	1.1041	1.1042	1.1041
	LT	2.8007min	2.8582	2.8955	2.9240	2.9479	2.9666	2.981[illegible]
	c_2, i_2	1.3, 3	1.3, 4	1.3, 4	1.4, 4	1.4, 4	1.4, 4	1.4, 4

Fig. 3 Pair map example of ellipse versus hyperbolic type: Job shop case

4.2 A Management Information Logic and Flow

Using 4.1, the objective formulation of IDTC-BG type would be similar to a management logic and flow at H = W body. Figure 5 shows the information flow, that is, another representation of Fig. 4.

Now, we could now summarize the management and information system of method and time-typed processes. From Figs. 1, 4 and 5, the following cyclic image of PDCA-CAPD versus IDTC-BG is a chain of DX (digital transformation) processes, and is seen at:

Fig. 4 Matsui's matrix equation of IDTC-BG type: An example of a convenience store

< lower level >

Ⓘ Ⓓ Ⓣ

$$\begin{pmatrix} d_t & \hat{d}_{t+1} \end{pmatrix} \begin{pmatrix} & \bar{O}_t \quad \hat{O}_{t+1} \\ d_t & \\ & (O\ d) \\ \hat{d}_{t+1} & \end{pmatrix} \begin{pmatrix} {}^{\exists}T < e_1\ e_2 > \\ \text{s.t. } D \to E \\ \text{by. } Z = Tx \end{pmatrix}$$

Ⓒ Ⓑ Ⓖ "compass"

$$\times \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} ER_t & LT_t \\ ER_{t+1} & LT_{t+1} \end{pmatrix} = \hat{g}(ER_{t+1}, LT_{t+1})$$

$\text{fund}: W = ER \times LT \to max$

< top level >

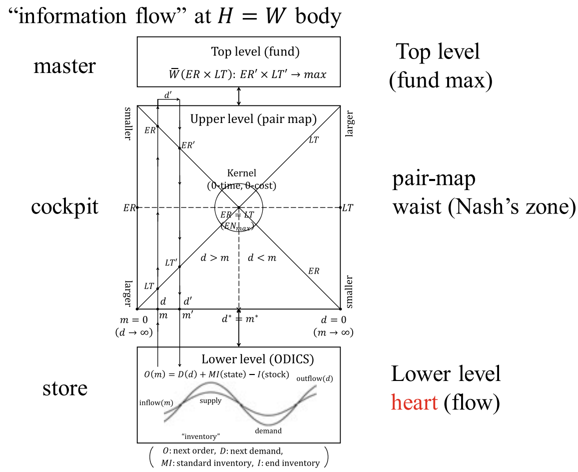


Fig. 5 Information flow of convenience store robot type at H = W body

$$\text{Mastsui's (Fig.4)}: \underset{\text{upper level}}{\overset{\text{lower level}}{}}\ \underset{\text{plan}}{\overset{\text{check}}{I(d, m)}} \rightarrow \underset{\text{do}}{\overset{\text{act}}{D(ER, LT)}} \rightarrow \underset{\text{check}}{\overset{\text{plan}}{T(d')}} \rightarrow \underset{\text{act}}{\overset{\text{do}}{C(ER'LT')}}$$

$$\rightarrow \underset{\text{(P)}}{\overset{\text{check}}{B(d, m')}} \rightarrow \underset{\text{(D)}}{\overset{\text{act}}{G}} \dashrightarrow \tag{9}$$

5 Conclusions

The development of an enterprise robot has been promoted since 2016 by the Research Institute for Engineering, Kanagawa University, Japan. Our first article in English was presented at ICPR2019 and published at Procedia Manufacturing 2020 [5]. Recently, the similarity of the clock system of body artifacts and circadian clock of biology is found in the nature versus artifacts study.

This second article focuses on the Sollen of body artifacts by developing the enterprise robot with a humanized body type and gives an applicability to the higher enterprise management of the new method and time managed-enterprise type. By pair-map, the modern enterprise could operate and manage its body at the middle of the clock scheme and Nash's zone in hyperboloid.

Our management would be similar to the "propagation-like" clock scheme of Schrödinger-type at quantum mechanism. There might be neither logics of artificial intelligence nor a human brain, and thus, there might be another managerial world and method in nature versus artifacts science.

References

1. Matsui, M.: Fundamentals and Principles of Artifacts Science: 3M&I-Body System. Springer Briefs in Business (2016)
2. Matsui, M.: Theory. Formulation and Realization of Artifacts Science: 3M&I-Body System. Springer Briefs in Business (2019)
3. Matsui, M.: Management of Manufacturing Enterprise: Profit Maximization and Factory Science. Kyouritsu-Shuppan, Tokyo (2005)
4. Matsui, M.: A game-theoretic consideration of order-selection and switch-over policy. Reprints of the Japan Industrial Management Association, Fall Meeting, pp. 48–49 (1983). (In Japanese)
5. Matsui, M., Ishii, N.: A demand-to-supply enterprise robot and its ODICS II type for convenience store application. Procedia Manuf. **39**, 381–388 (2019). https://doi.org/10.1016/j.promfg.2020.01.381
6. Taylor, F.W.: Scientific Management. Harper & Brothers, New York (1947)
7. Matsui, M.: A chameleon's harmonic correlation, criteria, and diversity for a win-win and sharing world. Theor. Econ. Lett. **10**, 735–744 (2020). https://doi.org/10.4236/tel.2020.104045
8. Matsui, M.: Manufacturing and Service Enterprise with Risks: A Stochastic Management Approach, International Series in OR&MS, vol. 125, Springer, Berlin (2008)
9. Markson, J.S., O'Shea, E.K.: The molecular clockwork of a protein-based circulation oscillator. FEBS Lett. **583**(24) 3938–3947 (2009). https://doi.org/10.1016/j.febslet.2009.11.021

Profiles and Testing in OPC UA—Current State and Future Potentials in Industry 4.0

Tonja Heinemann, Armin Lechler, and Oliver Riedel

Abstract We are moving towards a future production in which machines are networked with each other and offer new possibilities for process control and optimization. These goals are achievable with OPC UA, which is currently increasing in popularity. To achieve this connectivity, interoperability between the machines must be guaranteed. OPC UA provides concepts for testing to achieve interoperability. However, these have not yet reached their full potential. In addition to being used for certification with the OPC Foundation, these concepts may also improve the development of OPC UA products in the future as well as the integration of such products in factory environments. In this paper, the tools and information in the domain of OPC UA Profiles and testing that are publicly available, are outlined and critically evaluated. They are put in context with use cases of OPC UA developers and standardization working groups. The findings are compared to examples with similar requirements in another domain, Linux package managers. This comparison outlines possible improvements in the handling of OPC UA Profiles and test cases, which are explained in greater detail.

Keywords OPC UA · OPC UA profiles · Conformance units

1 Introduction

Open Platform Communications Unified Architecture (OPC UA) as a basis for communication is growing in importance in factory environments [1]. This can be attributed to the manufacturer independence and the use of well-established and well-known communication protocols used [2]. Another reason for the popularity of OPC UA is the option to use state of the art security features, which is not always possible for other means of shop floor communication [1, 2].

As the popularity of OPC UA rises, more and more so-called Companion Specifications (CS) are being created. These CS provide OPC UA data models, most

T. Heinemann (✉) · A. Lechler · O. Riedel
University of Stuttgart, ISW, Stuttgart, Germany
e-mail: tonja.heinemann@isw.uni-stuttgart.de

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_11

often for specific domains. Currently, more than 30 such CS exist. The number of published CS over the last years shows an upwards trend [3].

Parallel to the CS, the OPC UA base specification is also under constant development. This enables OPC UA to use advancements in all areas of communication technologies, like transport protocols, security mechanisms and data encoding. Especially in terms of security, this ability to update and upgrade the specification is important. This is due to existing encryption technologies being breached, and in response, becoming deprecated and replaced [4].

When OPC UA is implemented on a shop floor with production equipment from different domains, the OPC UA base specification and multiple CS are used together. In this scenario, the interoperability of all OPC UA devices is of great importance. This includes matching options to be implemented on all devices regarding transport protocols and encodings, as OPC UA allows for different options in this domain. It also includes matching versions of the different specifications needed [5]. On a single device, versions get especially important if more than one specification is implemented: if two or more specifications depend on different versions of a third specification, problems can occur.

There exists an OPC UA mechanism to keep an overview of implemented functions and versions: Profiles. Profiles bundle functionality on OPC UA Servers, may depend on other Profiles and are directly related to testing. When testing different OPC UA Products, the baseline requirements for interoperability can be ensured. Additionally, there is an option for OPC UA Servers to indicate which Profiles they support, namely the “ServerProfileArray” in the “AddressSpace” [5].

As the number of CS increases and with it the number of available Profiles, working with OPC UA requires more attention to detail in this domain. In this paper, different tasks of working with OPC UA are presented along with their requirements regarding Profiles and tests. The current tools available for OPC UA are presented and evaluated against those requirements. As version management is involved, package and dependency managers are presented. These programs are designed for dependency management of libraries and executable programs, a domain with a higher number of such dependencies than OPC UA has today. Based on a comparison of areas of use of these package managers and OPC UA Profiles, possible means to translate the version management capabilities to OPC UA are outlined.

2 Profiles and Test Cases in OPC UA

The IEC standard OPC UA aggregates technical solutions for multiple layers of industrial communication. This includes options for the transport layer and transport protocols, options for different encodings, all services and messages required for the client/server and publish/subscribe paradigm, the rules for representing data in models and the most basic elements to build such models on. As there are multiple options for any of those layers, both testing for correct implementation and an overview of supported options are important for successful communication. This is realised in OPC UA with Profiles, Conformance Units (CUs) and test cases [6].

Profiles are used to describe functionalities as a whole unit. They are subdivided in full featured Profiles, which contain options for all layers in an OPC UA application, and Facets, which focus on a functionality only in a subset of layers. All Profiles provide a “ProfileCategory”, which is used to group them in the specification documents and similar resources. They have a Uniform Resource Identifier (URI) to identify each Profile. All Profiles are groups of smaller units, which may be both Profiles and CUs. The exact list which Profiles and CUs are contained, is given in the Profile’s definition [5].

CUs are individually testable items; each CU is associated directly with test cases. In the specification documents, the CUs are defined with a ProfileCategory as well. In addition, they have a ConformanceGroup. Again, the ConformanceGroup is used for representation and organisation in the specification documents. As ConformanceUnits are individual entities, they can be used in more than one Profile [5].

The test cases for each CU are defined along with the specification, but not included in the specification release. OPC UA test cases may include both automatically executable tests as program code and manual tests with a description for the testing person [7]. These test cases are used to confirm conformance to OPC UA Specifications, in this case they are executed by an OPC Foundation accredited test lab [7].

For both Profiles and CUs, there may be changes when the specification they belong to is updated. If a Profile is changed in a new version of a specification, the year is added to the name, e.g. “Standard 2017 UA Server Profile” as the updated version of “Standard UA Server Profile”. For CUs, a new version is often indicated with a number appended to the name, e.g. “Address Space Method 2” as the updated version of “Address Space Method” [5, 8]. There is no property of Profiles or CUs individually indicating a version or revision.

CUs and Profiles for CS are defined by the same group that creates the specification. They are usually designed based on the domain model in the specification. When realizing OPC UA functionalities in a product, Profiles and CUs need to be regarded for certification. In the following section, the requirements of specification developers and product developers regarding Profiles are discussed. In Sect. 4, the tools and resources commonly used for these tasks are outlined.

3 Requirements of OPC UA Developers Related to Tests and Profiles

The view on tests and Profiles differs depending on whether they are defined in the specification, implemented in products or expected when connecting to an OPC UA product. Thus, this section is structured in three subsections highlighting the respective viewpoint. In this overview, the certification is only regarded for the data provider, as this is the more relevant case regarding CS. There are usually more test cases towards data providers, especially for domain models. For a data provider, it is important to implement domain models in the specified extent, as this is expected by data consumers. How much of this model is actually used by a data consumer is not important from the point of view of specification conformance. However, a data consumer can be certified as well, most importantly regarding the correct implementation of communication mechanisms [7].

3.1 Writing Companion Specifications

When writing a CS, a so-called Joint Working Group (JWG) consisting of domain experts is formed. This JWG defines the scope of the specification, the use cases it should support. The JWG gathers the requirements to represent these use cases on the OPC UA interface. With these prerequisites, the JWG designs a domain model following the modelling principles and rules of OPC UA. This model is then represented in the Companion Specification document, where additional textual parts describe and explain the model and its intended use. The JWG also designs the Profiles and CUs, and represents them in the same text document. When publishing the CS, the information model is additionally represented as a machine-readable XML file. To support conformance testing, the JWG can also define and implement test cases [9].

When defining Profiles, especially full featured Profiles, other Profiles and CUs from other specifications are often included. This is especially common with the full featured Profiles of the OPC UA base specification, describing OPC UA Servers and Publishers [5]. So, the JWG has to get an overview of existing Profiles and CUs and narrow the selection down to the contents needed for the CS.

When formulating the description for Profiles and CUs, it is advisable to keep descriptions similar to other specifications if the same basic concept is covered. CUs and Profiles in different CS are often similar, because they mainly concern model elements. The same applies to test cases: test cases concerning the correctness of model elements are often similar in structure. To achieve similarity to other specifications, the JWG needs to consider either those specifications/test cases or a secondary source providing a formulation/implementation template with the specific model context left blank [10].

After publishing a specification, JWGs may update and extend it as well as correct errors. One example is the specification OPC 10,030 UA for ISA-95, which is currently updated to include “Interfaces”, a concept that was introduced to OPC UA after the CS was published [11]. Another important reason for updating CS is the fact, that Profiles and CUs can be deprecated [5]. In this case, it would be beneficial to update the Profiles of any CS using the deprecated items. The prerequisite for JWGs to do this is to stay updated about changes in Profiles and CUs in other specifications.

3.2 Implementing and Certifying a Data Provider

When a developer implements an OPC UA data provider that conforms to one or more OPC UA CS, the conformance with these CS can be tested [5]. Ideally, the developer has access to the test cases developed by the JWG. In addition to testing with the goal of correct implementation, the product can also be certified for compliance. In this case, the product is tested in a test lab accredited by the OPC Foundation, which has always access to the test cases defined by the JWG [7].

The implementation, especially concerning the lower communication layers, can be simplified by using a Software Development Kit (SDK). In addition, a developer needs the information about required and available Profiles, as a prerequisite to implementing them. When implementing multiple CS in a product, the version information may be important. Especially, if there are dependencies on different versions of the same/similar Profiles or CUs.

3.3 Implementing a Data Consumer

Data consumers for OPC UA may vary a lot. The purpose of a data consumer may be a simple interface for few key values, as well as a process management system that is able to control the production via OPC UA. But regardless of the type of application, the OPC UA communication always follows the principles specified by the OPC Foundation. Regarding the data available to the consumer, CS help to define the data a consumer can expect from an OPC UA data provider. As such, they can help the developer to a certain extent.

However, many Profiles and CUs are optional for data providers, so the data consumer may need to handle similar providers with different feature sets. In this case, the consumer needs to check what Profiles and CUs are implemented by the data provider. This can either be done by consuming the information a data provider gives in this regard, the self-description of the data provider. The other option is to check, whether data providers support certain requirements of the data consumer application by testing.

The developer for the data consumer needs the information, what Profiles and CUs can be expected from data providers. In addition, the information regarding the implemented Profiles and CUs is important per data provider. To get information about the implemented Profiles/CUs, test cases can be used if available.

4 OPC UA Tools Available Today

This section lists the tools available for working with OPC UA today and puts them in context to the requirements of OPC UA developers outlined in the previous section.

Specification Documents and the CS Template
The specification documents for the base specification and the CS contain all information about Profiles and CUs. Cross-referencing and including Profiles/CUs from other specifications is possible. Test cases are not included in these documents. These documents are written by the OPC Foundation in the case of the base specification and the JWGs for the Companion Specifications, using the template provided by the OPC Foundation [12]. When developing products that adhere to the specifications, the developers use them as a source of information. The specifications are available publicly via the IEC, the OPC Foundation or the OPC UA Online Reference provided by the OPC Foundation [3, 6, 13, 14].

UANodeSet XML Files
The contents of information models can be represented in a machine-readable form, following a schema defined in [15]. For CS defining a model, such an XML document is usually provided with the specification. These files may be used by developers to aid in using the information model. Profiles and CUs are not included in these XML files. The information models of OPC UA Specifications are available online [16].

Modelling Tools
There are certain tools to help create and use OPC UA models. These tools can be used to create UANodeSet XML files or to use such files and create models for the use in products. In the model, there is a data element called "ServerProfileArray" that indicates the Profiles implemented by an OPC UA Server. This element can be filled using a modelling tool. Other than that, there is no direct connection to Profiles and CUs. These modelling tools are available mostly as commercial software, e.g. [17, 18].

OPC UA NodeSet Validator
The OPC UA NodeSet Validator application is used to check the identity of the model description in a CS document and the corresponding UANodeSet XML file. This tool is designed for JWGs to find inconsistencies between the documents quickly. It also checks formal elements of the specification document to ensure consistent specifications, even when written by different JWGs. The OPC UA NodeSet Validator is available online as a website [19].

OPC UA Compliance Test Tool and Test Cases
The OPC UA Compliance Test Tool (UACTT) is used for executing the automated tests for the CUs. It can be used by product developers to test their products as well as by certification test labs. The UACTT is available for OPC Foundation Corporate Members and includes all test cases as scripts for the tool [20]. Certification in a test

lab does not require an OPC Foundation membership. The test case descriptions are available for all OPC Foundation members [21, 22].

Profile Reporting Tool

The Profile reporting tool is a web resource providing an overview of Profiles and CUs defined by the OPC UA base specifications. It can be used by JWGs and developers to get an overview of Profiles and their dependencies. The Profiles and CUs of CS are not yet included; however, this is planned. The Profile reporting tool is available as a public web page [8].

Software Development Kits (SDKs)

SDKs are available for OPC UA Servers, Clients, Publishers and Subscribers. They provide OPC UA compliant communication and functions for all layers covered by the base specification, however, this may vary between SDKs. Some SDKs provide support for UANodeSet XML files or integration with modelling tools. SDKs are used by product developers, for data providers and data consumers alike. Regarding tests, Profiles and CUs, there is no formal connection, as with the modelling tools. SDKs themselves cannot be certified, but products build with an SDK can [7]. This way, an SDK can indicate that products compliant to the standard can be developed using it. Regarding availability, there are open source versions of SDKs as well as commercially available ones, e.g. [23–26].

As outlined in this section, there are different tools for different requirements. One recurring requirement for the specification development, as well as for the product development is a global overview of Profiles. For the JWGs developing the specifications, an overview is needed to initially define and continuously monitor depending Profiles and CUs. Concerning the modelling tools, a Profile overview could help while designing a model or a product and inform about deprecated Profiles and dependencies. The OPC UA NodeSet Validator could include checks for dependency problems and deprecations in terms of Profiles and CUs. In the case of the UACTT, a global overview of Profiles including version changes and deprecations could reduce the efforts needed to keep the UACTT up to date. With a global overview, all dependencies could be evaluated without the need to include them into the UACTT logic.

The closest resource that covers such requirements today is the “Profile reporting tool”. To date, it still lacks the information about CS. For JWGs and developers, a search functionality is missing. Even a simple search reduces the time to find a Profile or CU with a known name or URI and allows to find similar elements. At the moment, the Profile reporting tool is structured with a hierarchical view of Profiles on the left side. This leads to multiple appearances of Profiles, both as individual entry and as dependency. Such a representation is not immediately intuitive. For integration in tools like modelling tools, environments using SDKs or the Profile reporting tool, an application interface would be a prerequisite.

5 Dependency Management in Software Package Managers

Sections 3 and 4 show, there is an emerging need to handle Profile dependencies in OPC UA specifications. As reference, the dependency management of package manager programs is outlined in this section. They provide established methods to deal with dependencies across software packages. Differences and similarities to handling OPC UA Profiles are listed.

5.1 Software Package Mangers

Software package managers are commonly used to keep different pieces of software (“packages”) working together on a system. They are used to install and update libraries and programs on servers and personal computers, e.g. for different Linux distributions [27, 28]. Another major use are software library indices, like “pip” for the Python Package Index or “npm” for Node.js [29, 30]. These package managers have access to one or more databases containing the available packages and relevant metadata. From a client perspective, the package manager allows to query the available packages, to install and update software, and to uninstall software. During all of these operations, dependencies of the packages are regarded. On install, missing dependencies are also installed and dependencies already present on the system are checked especially regarding versions. If a dependency is already installed, it often is needed by another package as well. During uninstalling, there is the option to uninstall all dependencies that are not used by any other package as well. When updating packages, a package manager keeps track of the installed and required versions and replaces deprecated packages.

The information needed by package managers is provided by package developers with the package. This same information is represented in the database, both for the package managers to query and for users on a web interface like [31, 32]. Common package information is the name, version of the source code, release number for a program and os architecture the package is compiled for [33].

If a package depends on other packages, it lists these first level dependencies with name and version. The version can be given as a relation or a range (e.g. greater than version 3.2), as well as a single, fixed option. Dependencies of a package can be grouped in dependencies needed to build the software, dependencies needed to run the software and dependencies needed for the test suite. There may be optional dependencies that are not needed for the package to work but provide additional functionality. Packages may also list conflicting packages, that cannot be installed at the same time [33].

So, the information needed by the package managers is basic information about the package like name, version and information about the first level dependencies. Using this information, package managers can resolve deeper dependency hierarchies and install the most appropriate option for dependencies required by multiple packages.

5.2 Differences and Similarities Between Package Management and OPC UA Profile Management

The focus of package managers and managing OPC UA Profiles and test cases differs. While package managers are used to install and uninstall software, OPC UA Profiles describe a programmed functionality and provide tests to check this functionality. These checks are used in conformance certification for OPC UA products. With package managers, there is no focus on certification. Another notable difference is the much larger amount of software packages. The Profiles in the OPC UA environment are fewer in number, though continuously growing. For this reason, managing dependencies has long been important for package managers, but is only now becoming a necessity for OPC UA.

Although there are differences, package managers and managing OPC UA Profiles have some similarities as well. Both packages and OPC UA Profiles encapsulate functionality, though in different ways. Packages directly provide said functionality while Profiles describe it.

The most apparent similarity is that dependency management is a key factor in both domains. A common factor within the dependency management is to keep track of updated packages/Profiles and ensure backwards compatibility.

In both domains, the local and global views are important. The local view concerns which packages/Profiles are available on a system. The global view provides all available packages/Profiles with their properties and dependency information.

6 Using Dependency Management Information for OPC UA Profiles

As shown in Sect. 4, the usage and development of OPC UA Profiles could be improved by a Profile overview, that is searchable and provides and Application Programming Interface (API) to be accessed programmatically. In Sect. 5, it is shown, how such an overview helps in version handling by providing the basic information needed by package managers. As outlined in Sect. 5, the solutions for version handling in package managers are likely to be applicable to the domain of OPC UA Profiles as well.

A similar solution in the OPC UA domain could improve the overview of Profiles and their updates, deprecations and versions when developing specifications. Such a functionality could even be integrated in modelling tools and development environments, as long as the Profile overview provides an API for these programs. This is increasingly important, as updated specifications will likely deprecate some CUs and Profiles, e.g. the latest update of the base specification [34].

Another possibility based on a Profile overview is informing specification editors of changes in the Profiles and CUs they include from other specifications. With this information, they can update the specification accordingly. The use of deprecated Profiles could also be checked by the OPC UA “NodeSet Validator”.

When testing OPC UA Products for conformance, the test tools like the UACTT could use a global database to query dependency information. This reduces the development effort for the test tools, as the most recent Profiles do not have to be included in each tool but can be accessed via the global Profile overview.

To simplify the provision of such a global Profile database, Profile information similar to package information of the package managers could be used. If this information is provided in a machine-readable file, the database updates can be automated. The main information needed for a database providing version and dependency information are included in the OPC UA Profile model today. This includes the CUs and Profiles included in a Profile, and the name of the Profile.

A feature not yet common for OPC UA Profiles is versioning. So far, Profiles are updated by creating a new Profile incompatible to the old version. This makes backwards compatible updates of specifications much harder: If the specification defines a Profile depending on a now deprecated Profile X, this Profile has to be replaced by a new version (including the replacement of X). As seen with package managers, versions can be handled by providing the version of the Profile itself and the versions of all Profiles it includes. These dependencies can be given with relative versions (greater than version y, between versions y and z, …).

7 Conclusion and Outlook

As OPC UA gains popularity and more and more CS are developed, keeping an overview of possible Profiles and CUs is important in testing, specification development and product development alike. This is especially important with respect to dependencies of Profiles on other Profiles and CUs, both to ensure that all prerequisites are met in products and to keep specifications up to date as required Profiles change. The current tools do not provide much assistance in this regard.

When compared with dependency and version management in package managers, which are common for software packages, established solutions appear to be usable in the OPC UA domain.

With a global Profile overview, similar to current package indices and databases, tools for developing OPC UA models and products could integrate Profile information. As testing for CS is an increasingly relevant topic, an overview of the certified Profiles for product developers and for product customers is important. Further automation of OPC UA development concerning the specification, product development and the usage of products, is an area of future research.

As many CS are developed and use Profiles and CUs from other specifications in increasing numbers, version management becomes more important. Apart from the dependency concept used today, there are emerging cases in dependency handling that have to be discussed. One example is excluding dependencies, Profiles that cannot be used together. Another case are Profiles that provide the same functionality and can be used interchangeably. With packages on computer systems, the context of such excluding dependencies is the system itself. For OPC UA products however, Profiles for the information model are often defined in the context of a model or a namespace. This means, excluding dependencies could be implemented in the same OPC UA Product, as long as each dependency is exclusively used in a namespace.

An adjacent topic to a Profiles database is the content of the Profiles and CUs themselves. As comparisons of Profiles get easier with such a database, their comparability becomes more important. Especially for information models, many CUs seem to have a similar goal and structure, e.g. ensuring a type in the model is implemented correctly in a product. To make these kinds of Profiles and CUs easily understandable, their representation should be similar as well. Both a system to generate similar representations for Profiles/CUs and suggestions for their exact content are subject to further research.

References

1. Adolphs, P., Bedenbender, H., Dirzus, D., Ehlich, M., Epple, U., Hankel, M., Heidel, R., Hoffmeister, M., Huhle, H., Kärcher, B., Koziolek, H., Pichler, R., Pollmeier, S., Schewe, F., Walter, A., Waser, B., Wollschlager, M.: Reference Architecture Model Industrie 4.0 (RAMI4.0). Status Report (2015). Accessed 24 Feb 2020
2. Hoppe, S.: Standardisierte horizontale und vertikale Kommunikation. In: Hompel, M. ten, Vogel-Heuser, B., Bauernhansl, T. (eds.) Handbuch Industrie 4.0. Springer Reference Technik, 2nd edn., pp. 371–390. Springer Vieweg; ProQuest, Berlin, [Ann Arbor, Michigan] (2017)
3. OPC Foundation: OPC Unified Architecture Information Models. https://opcfoundation.org/developer-tools/specifications-opc-ua-information-models (2021). Accessed 17 Feb 2021
4. OPC Foundation: OPC 10000-2: OPC Unified Architecture. Part 2: Security Model, 1st edn. https://opcfoundation.org/developer-tools/specifications-unified-architecture/part-2-security-model/ (2018). Accessed 4 June 2021
5. OPC Foundation: OPC 10000-7: OPC Unified Architecture. Part 7: Profiles, 1st edn. https://opcfoundation.org/developer-tools/specifications-unified-architecture/part-7-profiles/ (2017). Accessed 4 June 2021
6. International Electrotechnical Commission: IEC 62541 (all parts): OPC Unified Architecture.
7. OPC Foundation: OPC Foundation. Part 1: Certification Process. Release 1.2 (2018)
8. OPC Foundation Unified Architecture Profile Reporting Tool. https://profiles.opcfoundation.org/v104/Reporting/index.htm?ModifyProfile.aspx?ProfileID=e711ce3a-efed-497d-bc45-02e452498729 (2016). Accessed 2 June 2021
9. OPC Foundation: Joint Working Groups—OPC Foundation. https://opcfoundation.org/about/working-groups/joint-working-groups/ (2020). Accessed 4 June 2021
10. Friedl, S., Heinemann, T.: Implicit Templates for Conformance Units in OPC UA Companion Specifications. 2021 IEEE International Conference on Industrial Technology, 2021. IEEE ICIT'21 (2021)
11. OPC Foundation <no-reply@opcfoundation.org>: OPC Foundation—Information—Edition 01 in 2021. E-Mail (2021)
12. OPC Foundation: Guidelines and Templates—OPC Foundation. https://opcfoundation.org/guidelines-and-templates/ (2021). Accessed 7 June 2021
13. OPC Foundation: Unified Architecture. OPC Unified Architecture Specification. https://opcfoundation.org/developer-tools/specifications-unified-architecture (2021). Accessed 7 June 2021
14. OPCFoundation: OPC UA Online Reference. Online versions of OPC UA specifications and information models. https://reference.opcfoundation.org/ (2021). Accessed 7 June 2021

15. OPC Foundation: OPC 10000–6: OPC Unified Architecture. Part 6: Mappings, 1st edn. https://opcfoundation.org/developer-tools/specifications-unified-architecture/part-6-Mappings (2017). Accessed 4 June 2021
16. https://github.com/OPCFoundation: OPC UA NodeSets and Other Supporting Files. https://github.com/OPCFoundation/UA-Nodeset (2021). Accessed 7 June 2021
17. Unified Automation GmbH: UaModeler "Turns Design into Code". https://www.unified-automation.com/products/development-tools/uamodeler.html (2021). Accessed 7 June 2021
18. Siemens Aktiengesellschaft: Siemens OPC UA Modeling Editor (SiOME) for implementing OPC UA companion specifications. https://support.industry.siemens.com/cs/document/109755133/siemens-opc-ua-modeling-editor-(siome)-for-implementing-opc-ua-companion-specifications?dti=0&lc=en-WW (2021). Accessed 7 June 2021
19. OPC UA NodeSet Validator. https://apps.opcfoundation.org/NodeSetValidator/ (2021). Accessed 2 June 2021
20. OPC Foundation: Test Tools. https://opcfoundation.org/developer-tools/certification-test-tools/opc-ua-compliance-test-tool-uactt/ (2021). Accessed 8 Feb 2021
21. OPC Foundation: Specifications. UA Client Test Cases. https://opcfoundation.org/developer-tools/certification-specifications/ua-client-test-cases/ (2021). Accessed 7 June 2021
22. OPC Foundation: Specifications. UA Server Test Cases. https://opcfoundation.org/developer-tools/certification-specifications/ua-server-test-cases/ (2021). Accessed 7 June 2021
23. Eclipse Milo. https://github.com/eclipse/milo. Accessed 7 June 2021
24. open62541. open62541 is an open source C (C99) implementation of OPC UA licensed under the Mozilla Public License v2.0. https://open62541.org/. Accessed 7 June 2021
25. Prosys OPC Ltd: OPC UA C/C++ SDKs. OPC UA C/C++ Client and Server SDKs are software development kits for speeding up your development in the OPC Unified Architecture world. https://www.prosysopc.com/products/opc-ua-cplusplus-sdk/. Accessed 7 June 2021
26. Unified Automation GmbH: Overview on Available SDKs. https://www.unified-automation.com/products/sdk-overview.html. Accessed 7 June 2021
27. Pacman Development Team: Pacman Home Page. https://archlinux.org/pacman/ (2021). Accessed 7 June 2021
28. Debian Wiki team: Apt—Debian Wiki. https://wiki.debian.org/Apt (2021). Accessed 7 June 2021
29. npm, Inc.: npm. https://www.npmjs.com/ (2021). Accessed 7 June 2021
30. PyPA: pip documentation v21.1.2. https://pip.pypa.io/en/stable/ (2021). Accessed 7 June 2021
31. PyPI: PyPI · The Python Package Index. https://pypi.org/ (2021). Accessed 7 June 2021
32. Arch Linux—Package Search. https://archlinux.org/packages/ (2021). Accessed 7 June 2021
33. PKGBUILD(5). https://archlinux.org/pacman/PKGBUILD.5.html (2021). Accessed 7 June 2021
34. OPC Foundation: OPC 10000: OPC Unified Architecture, 1st edn. https://opcfoundation.org/developer-tools/specifications-unified-architecture/specification-release-candidates-for-review/ (2021). Accessed 4 June 2021

Manufacturing Workflows in Microservice Architectures Supporting Digital Transactions for Business Process Automation

Jaime Garcia Represa and Jerker Delsing

Abstract Manufacturing systems are in the middle of a digital transformation. As systems in the assembly line are upgraded into cyber-physical systems (CPSs), capable of communicating between each other and carrying out complex computational tasks, the need for tight centralized control from an enterprise resource planning (ERP) and manufacturing execution system (MES) is less vital. In fact, not only manufacturing processes follow the trend toward decentralization and are moved to the edge layer. Other business processes along the supply chain have the potential to follow the digitalization process, such as procurement and supply flow management. This evolution brings new opportunities and challenges to the field. On the opportunity side, we identify shorter cycle times from product design to production, flexible production systems and multi-stakeholder production. Among the associated challenges, the collaboration of product, production, and business aware edge assets in multi-stakeholder environments stands out. This work proposes a new architecture for smart factories, in an environment where the products, supply stations and manufacturing equipment are controlled by different stakeholders. Requested manufacturing operations and supply flow are generated from machine-to-machine (M2M) negotiated business agreements between pairs of involved stakeholders. The manufacturing workflows are created and managed at each production workstation based on the smart product's needs. Operations and supply flow progress is logged in distributed ledgers for the involved pairs of stakeholders, providing non-repudiation and immutable data on the M2M business agreement. The proposed architecture enables the automation of business processes providing benefits in terms of decreased transaction time and cost.

Keywords Workflow management · Service oriented architecture · Smart factories · Arrowhead framework

J. Garcia Represa (✉) · J. Delsing
Luleå University of Technology, Luleå, Sweden
e-mail: jaime.garcia@ltu.se

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_12

1 Introduction

The manufacturing industry is in the midst of a digital revolution that is transforming its foundations. While the industry still holds an important stake in the GDP of developed countries, the benefits coming from pure manufacturing activity are predicted to stabilize, in contrast to the expected profit from manufacturing services [1]. Therefore, machinery manufacturers are no longer solely focusing on the manufacturing process performance of machines, instead computational power and digital communications have been identified as important factors to support service offerings.

These changes in business models, from a product-oriented perspective to a service-oriented model, have been promoted by several technological developments in the last decade in information technology (IT) and operational technology (OT), which are linked together through the Industry 4.0 initiative [2]. On a global scale, Industry 4.0 envisions connecting all companies involved in a supply chain, improving intercompany collaborations and information exchange through a shared business process. The observed trend points toward an increase in horizontal integration between companies' IT systems to reach end-to-end solutions. At the factory level, Industry 4.0 introduces the smart factory concept as an environment of connected and coordinated machines. The devices that will populate such factories are cyber-physical systems (CPSs), machines with a high degree of integration between the commanding software and acting hardware, their cyber and physical dimensions respectively. CPSs have their behavior tightly related amid both dimensions; therefore, changes in one part can affect the other, provided information can be shared between them through feedback loops [3].

In an effort to guide the development of Industry 4.0 solutions, a new set of architectures has been proposed. These architectures are published by different industrial consortia to meet their expectations of what the future industrial revolution should resemble by means of standardizing concepts and implementations. Among the most relevant architectures is the Reference Architecture Model for Industry 4.0 (RAMI 4.0) with its I4.0 component shell [4] that focuses primarily on the manufacturing industry, and the Industrial Internet Reference Architecture (IIRA) for Industrial Internet of Things (IIoT) [5], which is broader in scope as it tries to reach different industries: energy, healthcare, transport, etc. Together, they share key concepts such as the need for added connectivity between the organization's IT systems and shop-floor machines to facilitate new use cases related to predictive maintenance, autonomous operation, and distributed tracking (Fig. 1).

Traditionally, automation development efforts have focused on the assembly/manufacturing stage, while other business processes, such as supply chain management (SCM) or procurement, have been left mostly intact. However, the scope of Industry 4.0 is broader than previous industrial revolutions and it is expected to affect the whole value chain [6]. A key difference between manufacturing and procurement processes is the number of stakeholders involved. Manufacturing features more control by a single company over its processes, while in procurement

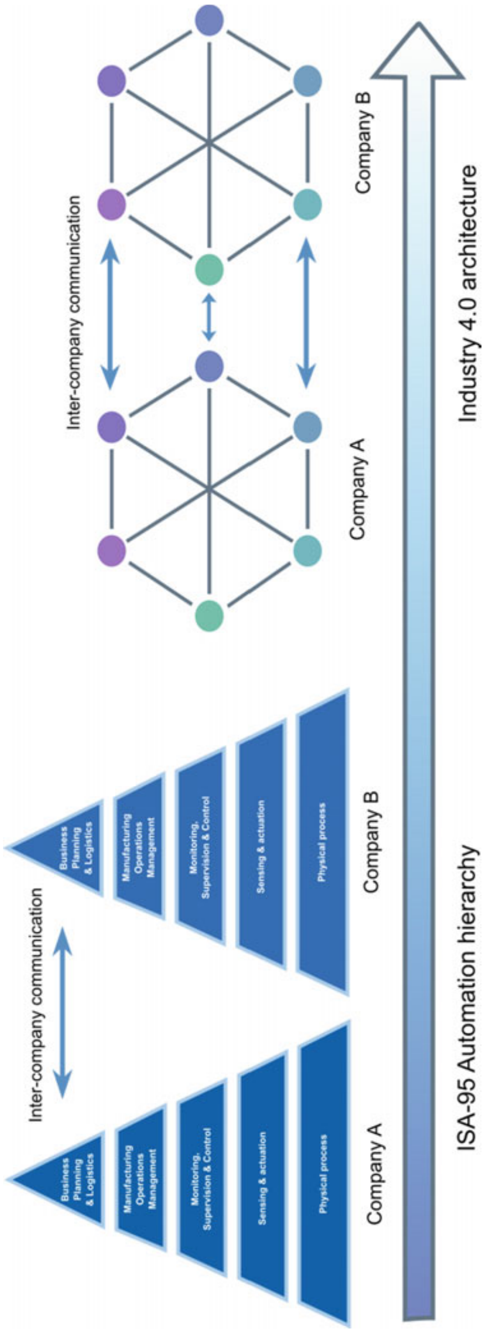


Fig. 1 Evolution of enterprise architectures

there is a collaboration between companies according to a commercial agreement. This creates a new set of challenges for which an automated and autonomous solution has not yet been found.

When analyzing common approaches to tackle the problem of complexity in engineering, whether it is due to the number of new CPS devices that need to be managed in a manufacturing facility or to the expansion in the number of stakeholders present at a global supply chain, modeling through workflows is a promising starting place. Engineers create models to abstract those details that are not useful to the problem at hand. Nevertheless, a static model of a factory or supply chain is not very useful when we expect to have a dynamic factory that supports flexibility and reconfiguration of manufacturing equipment. In addition, the goal of these new models is not only to support engineers in new designs but also to help manage and monitor plant operations. This opens the way to the use of workflows as a viable alternative due to their nature of allowing models with dynamic behavior and streamlined execution through automation.

The aim of this paper is to explore a new architecture for smart manufacturing through the use of workflows to integrate and automate shop-floor operations with supply chains and procurement processes. Guided by the multi-stakeholder environment foreseen in the fourth industrial revolution, we present a shop-floor scenario composed of smart products, a category of CPS with capabilities for machine-to-machine (M2M) communication with the devices at the manufacturing plant, as main actors of a series of business process triggering M2M agreements with the ultimate purpose of manufacturing each product as per its product recipe. The product recipe describes the manufacturing operations that should be performed upon the product, transferred to the smart product before it is placed in the factory assembly line.

The rest of this paper is organized as follows: Section II covers previous work on the topic of decentralized manufacturing systems. It also includes an introduction to the technologies used to develop our solution, covering workflows and business processes, microservice architectures and M2M agreements between CPSs. Section III describes the proposed manufacturing architecture in detail, with an example workflow to illustrate the service interactions. It reports on the concepts exposed during the example workflow showcasing its benefits. Section IV summarizes the contributions of our work and briefly discusses future directions of research.

2 Background and Supporting Technologies

In previous work [7], we proposed a transition from an ISA-95 [8] monolithic factory organization toward a distributed layout of production workstations. The main goal was to attain similar overall functionality at the manufacturing plant while migrating responsibilities for workflow control from a central manufacturing execution system (MES) to a set of distributed workstations. The approach proved beneficial to increase assembly line independence from higher-level IT systems, such as enterprise resource planning (ERP), product lifecycle management (PLM), and MES, and to improve its robustness and scalability. Moreover, it encouraged moving business logic to edge devices, reducing computer communications by having data processing closer to its source. However, our previous work was limited to the work performed at the assembly line and omitted supporting business processes.

In regard to a multi-stakeholder undertaking, there are numerous studies exploring different types of collaboration in manufacturing. Starting from the established ISA-95 enterprise systems, there has been research to expand their interfaces in [9], where workflows were used to model the interactions. We substitute centralized systems in favor of small, distributed nodes, but embrace modeling through workflows to control and track the communications between stakeholders. Wang [10] explores an agent approach to dynamic supply chains, where each software agent negotiates service compositions. The modeling of stakeholders as service requesters and service consumers is something that has been expanded upon in our work. However, we keep a simple model in negotiations without argumentation mainly based on iterative offers. In [11], the use of different workflow modeling notations has proven useful when working with heterogeneous business processes. This work does not focus on any specific workflow modeling language, leaving the decision open for the later implementation stage. Nevertheless, since there are workflows with different scopes and business processes, the option of their representation is not restricted to only one notation.

To explore the idea of an automated solution that expands various business processes, it was necessary to adopt several new technologies. In the following subsection, we will briefly describe their origins and evolution, as all the technologies will have a part in the proposed solution.

2.1 Microservice Architecture

Software architecture design is an ever-evolving field of computer science dealing with the high-level organization and structure of software systems, the main design decisions. As engineers change the way software components are deployed and connected or new requirements emerge from stakeholders, there is a need for new architectural patterns to capture the fundamental structures, features and context of such solutions. This allows for reuse of the architectural patterns in the future, whenever the same or similar problems are encountered.

Recently, the trend in software architecture has been to move away from monolithic solutions in favor of distributed applications. Accordingly, there was a rise in popularity of service-oriented architectures (SOAs) that proposed abstracting business logic into distributed services. In a SOA, the application logic is executed by communication between autonomous loosely coupled services, usually through lightweight protocols like HTTP or CoAP, as opposed to method invocations and function calls used in monolithic applications. However, the popularity of SOA played in its detriment as its definition broadened and increased in ambiguity.[1] From the pitfalls encountered with SOAs surged the microservice architecture style [12] with a refined and narrowed scope. This style emphasizes that intelligence should be encapsulated in the services, not in the communication channels, and that the size of services should be bounded, repurposing the Single Responsibility Principle of software design to service design.

The design according to a microservice architecture provides several benefits to an application. In the open source community, there are several frameworks supporting industrial requirements using a microservice or SOA [13]. For this work, the essential features needed are look-up, sometimes referred to as dynamic discovery, late binding, also called dynamic binding, and loose coupling. As analyzed in our previous work [7], the Eclipse Arrowhead framework meets those characteristics and is therefore the technology chosen for testing and prototyping our solution. It is a mature framework that has been tested in several European projects, such as Productive 4.0,[2] FAR-EDGE[3] and Arrowhead Tools,[4] together with a myriad of use cases ranging from machine tooling [14] to access control of time-series sensor data [15].

2.2 Workflows for Manufacturing

The study of workflows and business processes as a method to optimize procedures inside companies dates back to the first industrial revolution [16]. In this new industrial revolution, workflows are moved again to the forefront as a method to model and manage the interactions between CPS inside factories. The number of actions that can be automated and controlled through a workflow management system (WfMS) increases at the same rate as shop-floor machines are upgraded and become capable of communicating with IT systems. Before, tasks required machine resources that were indirectly commanded, a human operator was necessary to drive the machine. In a smart factory, shop-floor machines can directly receive commands and respond accordingly. However, this situation also carries new problems, in terms of interoperability between devices and software providers of WfMS.

Some authors in the field use the terms workflow and business process as synonyms [16], but in this work they are distinguished. First, we understand a business process as a set of tasks performed by resources in a specific order according to some predefined rules for the purpose of fulfilling a business goal. Through business processes, companies create value for their customers. Accordingly, workflow is defined in broader terms and used in other areas not necessarily tied to the business field. Workflows represent a similar set of tasks as business processes, but their objectives are not explicitly tied to a business goal. Therefore, it is possible to have workflows

[1] https://martinfowler.com/bliki/ServiceOrientedAmbiguity.html.

[2] https://productive40.eu/.

[3] http://www.faredge.eu/#/.

[4] https://www.arrowhead.eu/arrowheadtools.

that will fulfill several business processes or even none. In addition, the scientific community has expanded upon this notion of workflow and applied it to model scientific experiments in what has become known as scientific workflows [17] but such are not in the scope this work.

An emerging trend in the workflow management space is to use the service abstraction to command tasks of enterprise IT systems following a microservice or SOA. This type of architecture can also used at the lower levels to communicate with and between CPSs, streamlining the integration with WfMSs that are usually deployed at the higher levels.

2.3 M2M Negotiated Business Agreements

For the manufacturing process to be carried out successfully, it is required that the participants coordinate between themselves on the work to be performed by each stakeholder. This coordination requires communication between the participants in order to reach an agreement on what type of manufacturing operation to perform or what parts to use in the assembly of the product. Moving away from the established infrastructure, where such agreements come after negotiation between employees of each entity involved, we propose to fully automate these business processes by means of direct M2M communication between the CPS deployed along the shop-floor.

To manage and trust automated business agreements, the negotiation process and systems involved must comply with additional requirements that were previously implicit, as agreements were performed manually. The initial goal is to mirror how private industrial agreements between two parties are done today. Therefore, the three pillars of information security - confidentiality, integrity and availability - are well fitted, but not enough as they only cover exchanges of information not what they imply or who performed them. Additionally, the concepts of identification, authentication and non-repudiation are also essential to bind the agreement to a specific stakeholder. Further analysis of the confidentiality requirement revealed the need for negotiations without a central authority supervising and sanctioning the process. This requirement does not rule out the option to later expose the data when adjudication by a third-party authority is required. Altogether, they restrict the solution space until only a handful of technologies remain viable.

Instead of looking for an ad hoc solution or creating our own, a quick search in the scientific literature was performed. Similarities were found with work in the field of distributed ledger technologies and permissioned blockchains. Among the projects ready for industrial use, the Ethereum network and Hyperledger Fabric were among the most relevant for our purposes.

Additionally, inside the Productive 4.0 and Arrowhead Tools projects, there has been research conducted in this topic that is already integrated in a microservice architecture through the Eclipse Arrowhead framework and fulfills the requirements [18], so we limit our work to build-upon their solution. Therefore, in this work we will make use of the exchange network architecture for the negotiation of business agreements between CPSs. As a general purpose architecture, it should provide reusability in similar use cases. We chose the implementation based upon signature chains, due to its lack of governance, distributed data and better scalability [19]. The authors provide a prototype support system to test and redefine the specifications: the Contract Proxy (CoPx) system[5], which has been integrated in our solution.

3 Manufacturing Workflow Architecture

This section presents the main contribution of this work: a new architecture for smart manufacturing through the use of workflows and negotiated business agreements for factory shop-floor operations. Following the principles laid out by Industry 4.0, the architecture focuses on multi-stakeholder and intercompany collaboration within smart factories.

The factory model proposed in this work is the same as in our previous work [7], representing an assembly line divided into production workstations, also referred to as production cells, and a smart product winding through them. Production workstations and smart products are modeled as Arrowhead local clouds; thus, the software deployed inside them contains, in addition to the mandatory Arrowhead core systems, support systems to manage workflows, store information and negotiate business agreements. Physically, workstations are composed of manufacturing equipment and machine tools with a cyber interface as application systems. In the case of the smart product, the addition of commercial off-the-shelf electronics to the original product was the suggested upgrade, provided it did not have some computing power already installed.

As mentioned previously, our solution is aimed at an environment where several stakeholders are involved in the manufacturing process. These stakeholders can be broadly classified into three main groups: product owners, those who request the manufacturing of the products; manufacturers, the entities who own the manufacturing facility and the manufacturing equipment used to make products; and suppliers, those in charge of providing the raw materials and other parts needed to assemble the final products.

The mapping from stakeholders to components in the assembly line is straightforward. The product owner is responsible for the smart product, which he will tweak and configure according to his needs. The manufacturer controls the machine tools and equipment, which constitutes the production workstations in the assembly line. Finally, the suppliers manage the supply stations, or supply workstations. They can be modeled as workstations placed upstream in the material flow as they do not sit physically in the assembly line.

There are some preparations to be done before the smart product is to be placed in the assembly line. The product recipe has to be downloaded from ERP/PLM systems, in charge of demand planning and product development respectively, to the smart

[5] https://github.com/emanuelpalm/arrowhead-contract-proxy.

product internal storage. The product recipe is a file containing the specific features each manufactured product must have and is usually the result of a general product template personalized by the customer, in this case represented by the product owner. This approach to managing product information allows for batch-size one discrete production, an advantageous feature of Industry 4.0 production systems.

With the product recipe ready, the smart product can contact the MES to obtain an optimized pathway through the assembly line workstations. At the same time, the product will establish communication with the suppliers present at the plant to check for part availability and price. This set of conditions, part availability and price, could change according to the product owner needs, e.g., substitute parts or parts from a specific supplier could be chosen and the acceptable price range could be expanded according to market demand fluctuations. However, the topic of token exchange and price negotiation algorithms is out of the scope of this work. Once the conditions are met, the first business negotiations can begin between the product and its suppliers.

For the smart product to join the assembly line, it needs to be sure that whatever parts and supplies are needed for his first operations will be available by the time it reaches the first workstation. Therefore, the first M2M negotiated business agreements between the product and suppliers have to be concluded and executed beforehand. This order, of first requesting the material flow and then reaching the workstation that makes use of it, will be a constant requirement throughout the manufacturing process and repeated at each new step, workstation operation, along the assembly line. Analyzing communications at the system level, the process is similar regardless of the object under negotiation, with the main difference being the entities referred to as buyers and sellers. A step-by step description of the interactions is presented below (Fig. 2):

1. The process starts inside the buyer's local cloud with a web service consumer request from the buyer system (in this specific example, the role of buyer is played by the smart product) to the trusted contract negotiation service provided by the CoPx. Is important to recall that the CoPx is a support system implementing the Exchange Network architecture to assist in M2M negotiations by being capable of digitally signing and publishing contractual events. The request contains a proposal which asks for a certain number of parts delivered to certain workstation X at a specified time in exchange for a certain reward, e.g., future payment or value exchange.
2. The proposal will have a target identifier, supplier A, that is not part of the systems inside the smart product cloud. Therefore, the proposal will be transmitted through inter-cloud communications to peer local clouds containing their own CoPx system. The specifics of inter-cloud communications are omitted for clarity.
3. The receiving CoPx systems will, in turn, publish an event inside their local clouds for sellers listening to offers. This step follows a publish-subscribe messaging pattern enabled by the event handler support system, as the other system uses a request–response pattern.

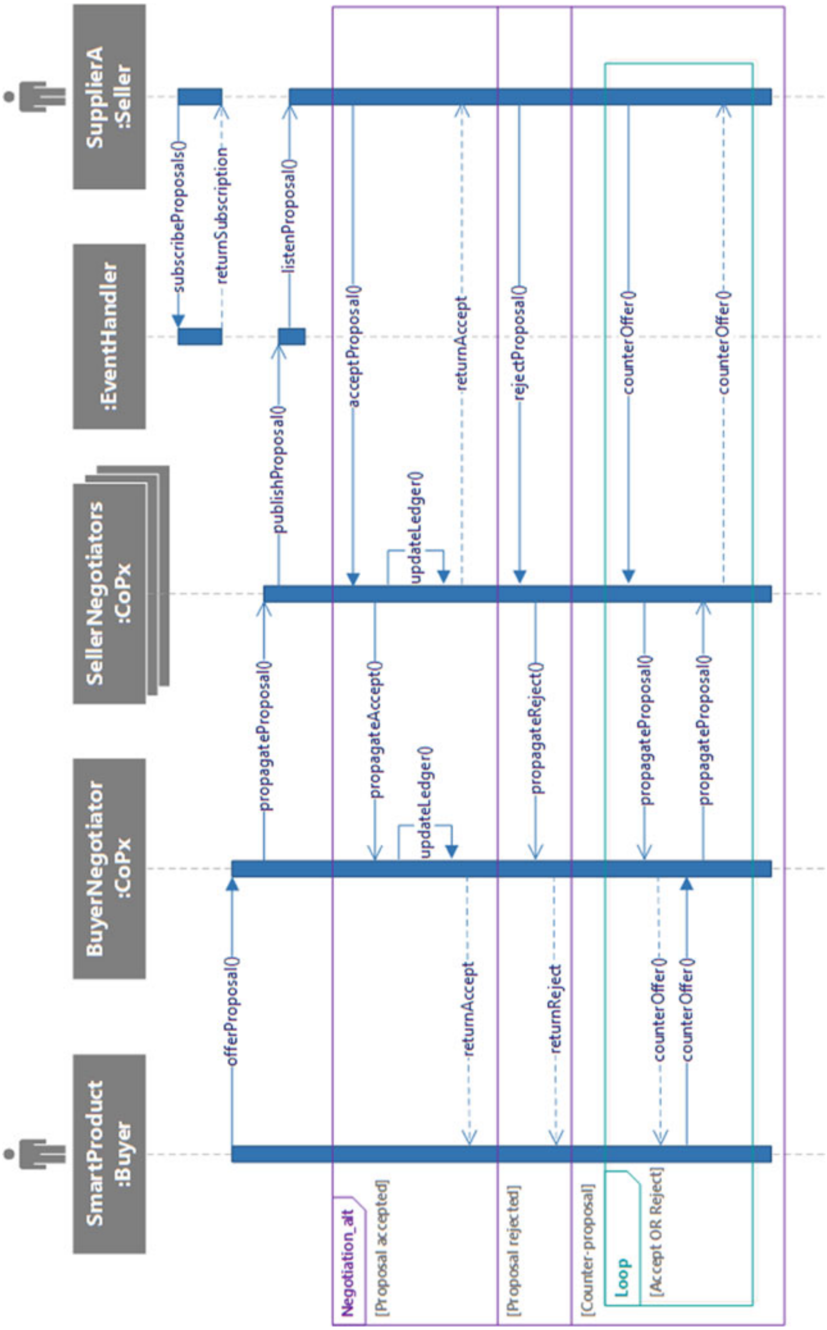


Fig. 2 Negotiation process modeled in UML sequence diagram

4. When the targeted seller system, which is supplier A in this example, receives an offer event with its ID it can reply to the requester in three ways. One way is by accepting the offer, another way is by making a counter-offer, and in the third case, it could reject the offer and close the negotiation. Whatever the chosen response is, it will return the response to the buyer through the network of CoPx systems, similar to how it arrived in step 2.
5. The negotiation is an iterative process that can take as long as necessary when considering new counter-offer and proposals. Eventually, when one of the entities accepts the deal, the proposal and agreement will be added to the exchange ledgers, based on signature chains, kept only by the entities involved in the negotiation as a trusted log of the business process.

Following a supply negotiation, the agreed parts will be shipped inside the factory to the indicated workstations. After the suppliers are notified of the delivery of the parts, the smart product signals its availability to join the assembly line, and the MES will proceed to its transportation into the first workstation. When it arrives at a workstation, there are multiple interactions at the system level between the smart product and the workstation local cloud systems (Fig. 3). These interactions span two main stages: the negotiation stage, where the goal is to arrive at an M2M negotiated business agreement on what manufacturing operations are to be performed, and the manufacturing stage, where the workflow manager and executor support systems request workstation manufacturing services. These service calls correspond to the task in a workflow created from the product recipe individually at each workstation. For an in-depth explanation of how the workflow is executed at the workstations, we refer to our previous work [7].

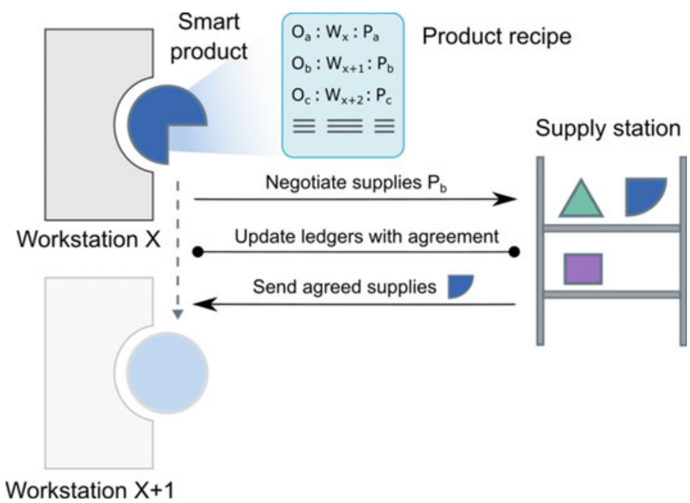


Fig. 3 Supply flow interactions between smart product and supplier according to product recipe

Summarizing the overall process, each smart product goes through a supplier negotiation stage, then a manufacturing negotiation step at the workstation, followed by its manufacturing according to the negotiated agreements. This process is repeated in subsequent rounds until the product reaches the end of the assembly line. The differences between one product and the next are mainly due to the specific information stored in the product recipe. Therefore, we identified the task pattern as a manufacturing workflow where the initial trigger is the product recipe.

4 Conclusions and Future Work

The digital evolution of the manufacturing industry is on its way. The equipment in manufacturing systems is being renovated to incorporate machines such as CPSs to adapt to and profit from the benefits promised by Industry 4.0 and its new service offerings. However, these new sources of revenue also entail an increase in the complexity of operations, for which automation cannot apply legacy solutions as manufacturing progresses from a single entity process to a multi-stakeholder venture.

The proposed architecture displaces the responsibilities of procurement and supply flow coordination from higher enterprise systems to CPSs deployed along the assembly line. There is a decentralization of functionality from large central systems such as ERP and MES, substituted by the use of smart products and M2M communications between any two CPSs, and supported through the abstraction of services in microservice architectures. Business processes increase their automation capabilities as devices can execute more business logic and are capable of negotiation without a supervising authority. They are able to reach agreements due to advances in distributed ledgers and blockchain technology. For designers to comprehend the interactions between all the systems involved in manufacturing and associated business processes, the architecture expands the use of workflows to model, capture and support the operations and negotiations.

A promising approach to decrease the design time and cost of Industry 4.0 solutions, like the architecture proposed in this paper, is a recent development of model-based engineering principles for manufacturing. The use of models in SysML that are always up-to-date and detail systems enough to support their automated implementation has potential [20].

Acknowledgements This research work has been funded by the European Commission, through the European H2020 research and innovation program, ECSEL Joint Undertaking, and National Funding Authorities from 18 involved countries under the research project Arrowhead Tools with Grant Agreement no. 826452. The authors would also like to thank the research project Productive 4.0 with Grant Agreement no. 737459.

References

1. Bedo, J., Strinati, E., Castellvi, S., Cherif, T., Frascolla, M., Haerick, W., Korthals, I., Lazaro, O., Sutedjo, E., Usatorre, L., et al.: 5G and the Factories of the Future. 5G-PPP White Paper (2015). https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Factories-of-the-Future-Vertical-Sector.pdf
2. Kagermann, H., Helbig, J., Hellinger, A., Wahlster, W.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Forschungsunion (2013)
3. Lee, E.A., Seshia, S.A.: Introduction to Embedded Systems: A Cyber-Physical Systems Approach, 2nd edn. MIT Press (2017). ISBN 978-0-262-53381-2
4. Adolphs, P., Bedenbender, H., Dirzus, D., Ehlich, M., Epple, U., Hankel, M., Heidel, R., Hoffmeister, M., Huhle, H., Kärcher, B., et al.: Reference Architecture Model Industrie 4.0 (RAMI4. 0). VDI-Verein Deutscher Ingenieure eV and ZVEI-German Electrical and Electronic Manufacturers Association, Status Report (2015)
5. Lin, S.W., Miller, B., Durand, J., Bleakley, G., Chigani, A., Martin, R., Murphy, B., Crawford, M., 2019. The Industrial Internet of Things, Volume G1: Reference architecture. Industrial Internet Consortium (2019). https://www.iiconsortium.org/pdf/IIRA-v1.9.pdf.
6. Hofmann, E., Rüsch, M.: Industry 4.0 and the current status as well as future prospects on logistics. Comput. Ind. **89**, 23–34 (2017). https://doi.org/10.1016/j.compind.2017.04.002
7. Garcia Represa, J., Delsing, J.: Autonomous production workstation operation, reconfiguration and synchronization. Procedia Manuf. **39**, 226–234 (2019). https://doi.org/10.1016/j.promfg.2020.01.329
8. Johnsson, C.: Isa 95—how and where can it be applied? ISA Expo 2004: Technical Conference And Emerging Technologies Conference (2004). ISBN 1-556-17895-6
9. Huang, C.Y.: Distributed manufacturing execution systems: a workflow perspective. J. Intell. Manuf. **13**, 485–497 (2002). https://doi.org/10.1023/A:1021097912698
10. Wang, M., Wang, H., Vogel, D., Kumar, K., Chiu, D.K.: Agent-based negotiation and decision making for dynamic supply chain formation. Eng. Appl. Artif. Intell. **22**, 1046–1055 (2009)
11. Kozma, D., Varga, P., Larrinaga, F.: Data-driven workflow management by utilising BPMN and CPN in IIoT systems with the Arrowhead framework. In: 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, pp. 385–392 (2019)
12. Newman, S.: Building Microservices: Designing Fine-Grained Systems. O'Reilly Media, Inc. (2015)
13. Paniagua, C., Delsing, J.: Industrial frameworks for internet of things: a survey. IEEE Syst. J. **15**(1), 1149–1159 (2020). https://doi.org/10.1109/JSYST.2020.2993323
14. Aldalur, I., Illarramendi, M., Larrinaga, F., Perez, T., Sáenz, F., Unamuno, G., Lazkanoiturburu, I.: Advantages of arrowhead framework for the machine tooling industry. In: IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, IEEE, pp. 4511–4518 (2020)
15. Chiquito, A., Bodin, U., Schelén, O.: Access control model for time series databases using ngac. In: 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, pp. 1001–1004 (2020)
16. Van Der Aalst, W., Van Hee, K.M., van Hee, K.: Workflow Management: Models, Methods, and Systems. MIT Press (2004)
17. Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., Lee, E.A., Tao, J., Zhao, Y.: Scientific workflow management and the kepler system. Concurr. Comput. Prac. Experience **18**, 1039–1065 (2006)
18. Palm, E., Bodin, U., Schelén, O.: Approaching non-disruptive distributed ledger technologies via the exchange network architecture. IEEE Access **8**, 12379–12393 (2020)
19. Palm, E., Schelén, O., Bodin, U., Hedman, R.: The exchange network: an architecture for the negotiation of non-repudiable token exchanges. In: IEEE 17th International Conference on Industrial Informatics (INDIN), IEEE, pp. 760–767 (2019)

20. Urgese, G., Azzoni, P., van Deventer, J., Delsing, J., Macii, E.: An engineering process model for managing a digitalised life-cycle of products in the industry 4.0. In: NOMS 2020–2020 IEEE/IFIP Network Operations and Management Symposium, IEEE, pp. 1–6 (2020)

Scheduling 3D Printing Machines to Minimize Makespan

Zachary Hassan, Gursel A. Suer, Jesus Pagan, You Yuqiu, and Neil Littell

Abstract Manufacturing industry has been evolving during the last few centuries. Industry 1.0 started with mechanization and the use of steam power. Mass production using production lines and assembly lines dominated Industry 2.0 era. Industry 3.0 era brought automation, flexibility and product diversity and Flexible Manufacturing Systems (FMS) and cellular systems were extensively used. Recently, there is a shift towards the fourth industrial revolution (Industry 4.0). Industry 4.0 includes the combination of technologies working together to fulfill a manufacturing task. Industry 4.0 utilizes internet of things (IIoT), big data, cloud computing, cybersecurity, autonomous robotics, augmented reality, and additive manufacturing (AM). The purpose of Industry 4.0 is to integrate the entire network to function as one system. In this study, we are focusing on scheduling 3D printing machines, namely Markforged Mark Two printers. Process parameters that can be considered in these printers are layer height, infill density, print speed, build orientation, infill patterns, and print temperature. These machines are Fused Filament Fabrication (FFF) 3D printers. The parameters considered in this study are infill density and layer height. Infill density dictates the amount of material that is filled on the inside of an object while it prints. Infill density has a role in a part's strength and weight. Generally speaking, the greater the infill density, the stronger and heavier an object will be.

Z. Hassan · G. A. Suer
Department of Industrial and Systems Engineering, Ohio University, 45701 Athens, OH, Greece
e-mail: zh995312@ohio.edu

G. A. Suer
e-mail: suer@ohio.edu

J. Pagan · Y. Yuqiu (✉) · N. Littell
Department of Engineering Technology and Management, Ohio University, 45701 Athens, OH, Greece
e-mail: youy@ohio.edu

J. Pagan
e-mail: paganj@ohio.edu

N. Littell
e-mail: littellw@ohio.edu

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_13

Lower infill densities on a part suggest that the object's intentions are purely visual with higher infill densities meant for functional parts. Markforged Mark Two allows infill density for rectangular infill to be from 0–92%. On the other hand, layer height determines the amount of material that is extruded through the nozzle during each pass. Markforged Mark Two allows for three different layer heights to be examined, 100, 125 and 200 mm. Layer height plays a large role in print time as the amount of material extruded effects the completion rate of the object. Layer height's impact can also be seen by a part's fineness or detail. This is represented visually on the object by being able to see each pass of the plastic material. For example, an object with a larger layer height will look rougher and not as smooth as an object with a lower layer height. However, it is well known that a lower layer height increases print time whereas a larger layer height implies a faster print time. Several parts with different geometries and also sizes are included in the study. The scheduling performance measure considered is makespan. The objective of the study is to find the optimal parameter settings for multiple jobs such that makespan is minimized subject to minimum restrictions on print parameters for various jobs. A mathematical model is presented to minimize makespan first. Once the optimal makespan is found, the model is re-run such that better quality parameter settings are determined while keeping the optimal makespan unchanged. Later, the results of the experimentation with various parts are discussed and future work is recommended.

Keywords 3D Printing · Scheduling · Makespan

1 Introduction

In this section, additive manufacturing considered in this paper is discussed in several subsections.

1.1 Additive Manufacturing (AM) and Fused Filament Fabrication (FFF)

AM begins with [1] a model on a computer aided design (CAD) software that dictates the desired geometry for a part. The file for that object is then converted to a stereolithographic (STL) file where a slicing software transforms the STL file into G-Code. The G-Code represents the toolpath the 3D-printer follows during its production time. Material [2], usually wrapped around a spool, is fed through a guide tube up to the print head nozzle. The material is then heated up to its melting point and dispenses material layer upon layer on a print bed by moving in the x, y, and z directions. Fused Deposition Modeling (FDM) is a material extrusion AM process. FDM [3] produces parts by heating filaments to its melting point, extruding the material through a nozzle, and laying filament on a print surface layer by layer until there is a completed part.

FFF is a type of FDM AM technology that uses carbon fiber reinforcement (CFR) while printing its parts. This paper will focus its study on an FFF printer.

1.2 Additive Manufacturing (AM) Versus Traditional Manufacturing (TM)

AM builds material layer by layer, whereas TM removes material from raw stock. Pereira [4] states how TM focuses on creating the highest amount of volume for the cheapest cost, whereas AM is meant for lower production and higher complexity parts. Due to cost models, Pereira concluded that TM is best for higher production volumes and AM is more appropriated when dealing with complexity or customization.

Jimo [5] states that main differences in AM processes versus TM starts with tool elimination as AM does not need a tool to operate, but TM does. This results in smaller batch production and a reduction in leadtime. Material efficiency is handled well because TM removes material and creates more waste whereas AM adds material and can use recyclable material. AM uses less inventory space allowing for freed up working capital to tend to other aspects of production. Lastly, compared to TM, AM reduces the number of changeovers, setups, and workers.

1.3 Advantages and Disadvantages of AM

Rylands [6] concludes that AM is beneficial from its reduced waste, customizable designs with little tooling, rapid prototyping, shorter lead times, lower inventories, and its ability to react and produce at the point of demand. On the other hand, some flaws stated included the training required to work with different materials and any software, low quality part finishes, slow printing speeds and narrowed capability.

Bas [7] discusses forty-three different problems FFF printers are currently experiencing; however, it should be noted that several failures found can be reflected from operator setup such as the nozzle being too close to the print bed and other settings that the operator decides on. The print failures include extrusion interruption, surface deformities, dimensional errors, oozing, surface scratches, overheating, warping, layer issues, and others. Garcia [8] continues the discussion by addressing FFF's simplicity, applauding its flexibility and versatility to manufacture complex shapes and geometries while creating minimal waste. On the other hand, Garcia states FFF's shortcomings in part quality noting its lower mechanical properties and geometric accuracy.

1.4 Objectives of the Study

The estimated print times are used to minimize makespan in a parallel 3D-printers shop. Makespan is defined as C_{max} in a multi-machine environment. The mathematical model chooses the best infill density parameter and layer height parameter to minimize makespan. Once the optimal makespan is determined, another model is used to find the higher quality parameters without violating the optimal makespan found in the earlier phase.

2 Literature Review

FFF specific literature either includes a focus on the physical properties of a material upon the application of forces or mechanical properties, the closeness of an object's measurement to its true value or geometric accuracy, and specific measures referring to the status of the process or process parameters. General AM literature discusses scheduling, part performance, or optimization models. Infill pattern is sporadically divided amongst the literature.

There are few scheduling problems that are discussed specifically for FFF printers, however general AM scheduling problems provide an insight to how the process works. Zhang [9] examined a parallel scheduling problem in regard to additive manufacturing by proposing an evolutionary algorithm. The study focused on a parallel 3D printed scheduling problem (P3DPSP). The aim was to optimize task assignments in a shop with multiple 3D printers. The overall objective was to minimise the maximum completion time of all batches.

Kucukkoc [10] developed mixed integer linear program models for additive manufacturing scheduling problems to minimize makespan. This scheduling problem analyzed various machine type considerations to develop a mathematical model for single and multiple machines to optimize makespan. The idea was to apply a scheduling problem to a conceptual AM production model. This consisted of a quantity of distributed parts, the grouping of parts, the distribution of those grouped parts to an AM machine, and the creation of jobs for machines.

Kim [11] used a scheduling method to establish the shortest production time of damaged or lost weapon parts from the use of multiple or parallel 3D printing machines. The study proposed a mathematical model to optimize production and confirmed that makespan can be shortened through the introduction of the subpart-split manufacturing concept for weapon parts. Subpart-split manufacturing is the idea of manufacturing different aspects of the same sub-assembly or part on more than one machine.

Fera [12] proposed a modified genetic algorithm for a single machine AM scheduling problem where they aimed to optimize time and cost in line with lean manufacturing principles. The paper analyzed how to schedule production orders for a single Direct Metal Laser Sintering (DMLS) AM machine. The AM machine scheduling problem (AMSP) used a multi-objective function subject to geometric volume and other variables. The combinatorial Optimization Problem (MOP) was divided based on the time and cost of parts. The aim was to find the best combination of jobs and quantities to be produced ahead of time.

Medina-Sanchez [13] provided insight to how average print speed can impact the estimation time for builds. Although it is not specifically related to scheduling, the ability to accurately predict build time is crucial for the scheduling process. The proposed linear-power (LP) and spline of Coons patches (CP) estimation methods were applied to three different FFF printers. The results examined absolute and relative errors compared to pronterface prediction and theoretical time. It was concluded that the LP and CP approaches were able to calculate a maximum relative error of 8.5% and speed approximations with a mean absolute error less than 2.7 mm/s. The ability to estimate build time is an essential tool for AM scheduling problems.

3 Problem Statement

The accessible Markforged Mark Two printer in the lab is capable of studying process parameters such as layer height, infill density, print speed, build orientation, infill patterns, and print temperature. Composites will already be considered because the machine is a Fused Filament Fabrication (FFF) 3D printer. Based on the information provided in the literature, this paper will use a rectangular infill pattern to study the effects of infill density and layer height on print time. The paper later proposes two mathematical models to minimize makespan to choose the infill density parameter and layer thickness parameter for each part. The shop configuration is identical 3D printer shop.

3.1 Product Breakdown

For the purpose of this study, a poker chip, chess pawn, and game die will be printed using a different combination of parameters which are shown in Fig. 1. The parameters will include a rectangular infill pattern, and varying amounts of layer height and infill density.

3.2 Infill Patterns

Infill patterns are used to determine how the inner structure of your part is created. Generally, infill is placed to optimize print quality and part strength. It is comprised of shapes that are distributed throughout each layer of the print. The Markforged Two utilizes three different patterns which consist of triangular, rectangular, and

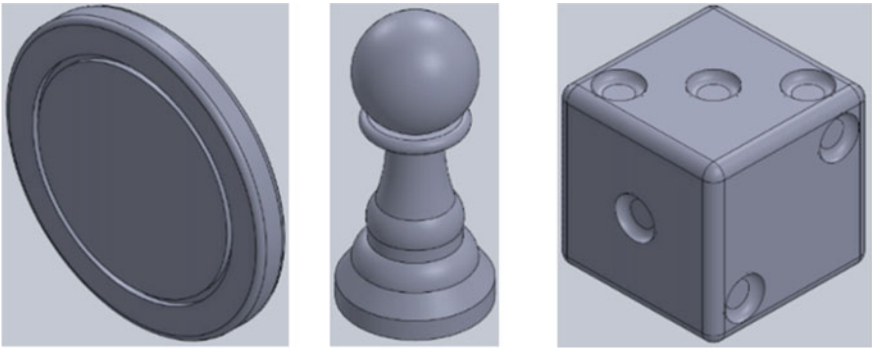


Fig. 1 Parts considered in this study

hexagonal. There is also a solid fill option as well, however with solid fill the entire inner structure is filled. Triangular fill is known to be one of the best infill options due to its support structure, strength, lack of deformity, and relatively quick printing speed. Rectangular infill is unique in that it is the only infill that can produce a part that is fully dense thanks to its grid of parallel and perpendicular extrusions. Since the print head moves in straight lines and does not have to readjust its position as often, the print speed is considered to be quick. Unlike triangular and rectangular, hexagonal infill is the slowest print speed because the print head must constantly be changing direction. However, it also produces the highest strength-to-weight ratio of any infill pattern. Since this paper will look at minimizing makespan, it is safe to say that hexagonal infill will most likely not be in contention for best printing method due to its longer print time.

The infill pattern for this study will be fixed with a rectangular infill. Rectangular infill has a range from 0–92% allowing the study to have access to the widest range of infill density. Rectangular infill is unique in that it is the only infill that can produce a part that is fully dense thanks to its grid of parallel and perpendicular extrusions. Since the print head moves in straight lines and does not have to readjust its position as often, the print speed is considered to be quick.

3.3 Layer Height

Layer height determines the amount of material that is extruded through the nozzle during each pass. Eiger.io allows for three different layer heights to be examined, 0.100, 0.125 and 0.200 mm. However, if carbon fiber reinforcement (CFR) is used, layer height is restricted to 0.125 mm. Layer height plays a large role in print time as the amount of material extruded effects the completion rate of the object. Layer height's impact can also be seen by a part's fineness or detail. This is represented visually on the object by being able to see each pass of the plastic material. For example, an object with a larger layer height will look rougher and not as smooth as an object with a lower layer height. However, it is well known that a lower layer height increases print time whereas a larger layer height implies a faster print time.

3.4 Infill Density

Infill density dictates the amount of material that is filled on the inside of an object while it prints. Infill density has a role in a part's strength and weight. Generally speaking, the greater the infill density, the stronger and heavier an object will be. Lower infill densities on a part suggest that the object's intentions are purely visual with higher infill densities meant for functional parts. Eiger.io allows infill density for rectangular infill to be from 0–92%. The larger range for infill density is used to determine the impact density has on print time. Infill density normally plays a role in studies that look at the mechanical properties of an object more so than a scheduling problem.

4 Methodology

4.1 Estimated Print Times

This experiment will be operated using the Markforged Mark Two 3D printer. Onyx, a nylon mixed with sliced carbon fiber, will be the primary material used. The study will include two parts. The first will analyze specific factors that affect print time and the second will conduct how the first part relates to scheduling. The factors to be studied include layer height and infill density with a fixed rectangular infill pattern for the parts. Layer height can be 0.1, 0.125 or 0.2 mm when no reinforcement material is used. The Eiger software controls this aspect. The Eiger software allows an infill density of 0–92% for rectangular fill. For this reason, the study will examine three different infill density percentages. To cover the range evenly, the parts will be printed used 10, 51 and 92% infill density. Three parts of three different sizes will be used to conduct the study consisting of a poker chip, chess pawn, and a game die. The various sizes and shaped objects allow for the study to comprise an overall better understanding regarding the experiment.

4.2 Scheduling Applications

Below are the two mathematical models presented in this paper. The first model minimizes the makespan. Once the optimal makespan is determined, the second model improves quality of the parts subject to the optimal makespan determined with the first model. The quality of parts improves by selecting lower layer thickness and higher infill density values.

Model 1:

Decision Variables:

MS Makespan.

X_{ijkm} 1 if job i with infill pattern density j and layer thickness k is assigned to machine m, 0 Otherwise.

Parameters:

P_{ijk} print time for part i with infill density j and layer thickness k.

n number of jobs.

f number of infill density options.

k number of layer thickness options.

m number of machines.

Indices:

i job index.

j infill density index.

t layer thickness index.

r machine index.

Objective Function:

$$\text{Min z} = \text{MS} \quad (1)$$

Subject to:

$$MS - \sum_{i=1}^{n}\sum_{j=1}^{f}\sum_{t=1}^{k} X_{ijtr} p_{ijt} \geq 0 \quad r = 1, 2, 3, \ldots, m \quad (2)$$

$$\sum_{j=1}^{f}\sum_{t=1}^{k}\sum_{r=1}^{m} X_{ijtr} = 1 \quad i = 1, 2, 3, \ldots, n \quad (3)$$

The objective function minimizes makespan as shown in Eq. 1. Equation 2 guarantees that makespan is greater than or equal to the completion of the last job on each machine. Finally, Eq. 3 ensures that each job is assigned to a machine. Of course, an infill density and layer thickness value are selected for each job as well.

Model 2:

Additional Decision Variable:

MS_r C_{max} for machine r.

Additional Parameter:

MS* The optimal makepan value determined in Model 1.

Objective Function:

$$\text{Max z} = \sum_{r=1}^{m} MS_r \tag{4}$$

Subject to:

$$MS_r - \sum_{i=1}^{n} \sum_{j=1}^{f} \sum_{t=1}^{k} X_{ijtr} p_{ijt} = 0 \quad r = 1, 2, 3, \ldots, m \tag{5}$$

$$\sum_{j=1}^{f} \sum_{t=1}^{k} \sum_{r=1}^{m} X_{ijtr} = 1 \quad i = 1, 2, 3, \ldots, n \tag{6}$$

$$MS_r \leq MS^* \tag{7}$$

The objective function of model 2 maximizes the completion times of jobs on all machines as shown in Eq. 4. Eq. 5 captures the maximum completion time on each machine. Equation 6 guarantees that each job is assigned to a machine. Finally, the optimal makespan value determined by the 1st model cannot be violated as given in Eq. 7.

5 An Example Problem and Results

In this example problem, there are six jobs, two infill density options (ID1, ID2) and two different layer thickness values (LT1, LT2) and two identical printers. The data for these four options is presented in Table 1. The gantt charts for Model 1 and Model 2 are presented in Figs. 2 and 3, respectively. The summary of job assignments for both models is also presented in Table 1. As one can see, two of jobs (jobs 1 and 2) will be printed using higher quality settings without violating the original optimal makespan.

Table 1 Data for the example problem and the solutions by model 1 and model 2

Option number	Options	Processing times						Makespan
		Job1	Job2	Job3	Job4	Job5	Job6	
1	ID1-LT1	13	19	22	14	34	28	
2	ID1-LT2	16	28	27	18	44	40	
3	ID2-LT1	15	20	26	15	43	30	
4	ID2-LT2	20	29	35	20	59	44	
Model 1 solution		M2 Op1	M2 Op1	M2 Op1	M2 Op1	M1 Op1	M1 Op1	MS1 = 62 MS2 = 68 MS = 68
Model 2 solution		M2 Op2	M1 Op3	M2 Op1	M1 Op1	M1 Op1	M2 Op1	MS1 = 68 MS2 = 66 MS = 68

Model Solution 1				
M1	Job 5	Job 6		
Processing Time	34	28		
Completion Time	34	62		
M2	Job 1	Job 2	Job 3	Job 4
Processing Time	13	19	22	14
Completion Time	13	32	54	68

Fig. 2 Gantt chart for model 1 solution

Model Solution 2			
M1	Job 2	Job 3	Job 5
Processing Time	20	14	34
Completion Time	20	34	68
M2	Job 1	Job 3	Job 6
Processing Time	16	22	28
Completion Time	16	38	66

Fig. 3 Gantt chart for model 2 solution

6 Future Work

This work is currently being extended to include more jobs, machines and options. We are also planning to modify the model to include minimum quality parameters established by the customers as an input to the model. Of course, with better quality parts comes cost considerations. These models will be modified to incorporate cost considerations as well.

References

1. Vorkapic, N., Pjevic, M., Popovic, M., Slavkovic, N., Zivanovic, S.: An additive manufacturing benchmark artifact and deviation measurement method. J. Mech. Sci. Technol. **34**(7), 3015–3026 (2020)
2. Thomas, D.S., Gilbert, S.W.: Costs and cost effectiveness of additive manufacturing. NIST Spec. Publ. **1176**, 1–7 (2014)
3. Alafaghani, A., Qattawi, A.: Investigating the effect of fused deposition modeling processing parameters using Taguchi design of experiment method. J. Manuf. Process. **36**, 164–174 (2018)
4. Pereira, T., Kennedy, J.V., Potgieter, J.: A comparison of traditional manufacturing vs additive manufacturing, the best method for the job. Proc. Manuf. **30**, 11–18 (2019)
5. Jimo, A., Braziotis, C., Rogers, H., Pawar, K.: Traditional vs Additive Manufacturing Supply Chain Configurations: A Comparative Case Study. Proc. Manuf. **39**, 765–774 (2019)
6. Rylands, B., Bohme, T., Gorkin, R., III., Fan, J., Birtchnell, T.: The adoption process and impact of additive manufacturing on manufacturing systems. J. Manuf. Technol. Manage. **27**(7), 969–989 (2016)
7. Bas, H., Elevli, S., Yapici, F.: Fault tree analysis for fused filament fabrication type three-dimensional printers. ASM Int. **19**, 1389–1400 (2019)
8. Garcia, E., Nunez, P.J., Chacon, J.M., Caminero, M.A., Kamarthi, S.: Comparative study of geometric properties of unreinforced PLA and PLA-Graphene composite materials applied to additive manufacturing using FFF technology. Polym. Testing **91**, 1–16 (2020)
9. Zhang, J., Yao, X., Li, Y.: Improved evolutionary algorithm for parallel batch processing machine scheduling in additive manufacturing. Int. J. Prod. Res. **58**(8), 2263–2282 (2019)
10. Kucukkoc, I.: MILP models to minimise makespan in additive manufacturing machine scheduling problems. Comput. Operat. Res. **105**, 58–67 (2019)
11. Kim, S.C., Kim, M., Ahn, N.: 3D Printer Scheduling for Shortest Time Production of Weapon Parts. Proc. Manuf. **39**, 439–446 (2019)
12. Fera, M., Fruggiero, F., Lambiase, A., Macchiaroli, R., Todisco, V.: A modified genetic algorithm for time and cost optimization of an additive manufacturing single-machine scheduling. Int. J. Ind. Eng. Comput. **9**, 423–438 (2018)
13. Medina-Sanchez, G., Dorado-Vicente, R., Torres-Jimenez, E., Lopez-Garcia, R.: Build time estimation for fused filament fabrication via average printing speed. MDPI Mater. **12**, 1–16 (2019)

Manufacturing: System and Automation

Electrode Tab Deflection Detection for Pouch Lithium-Ion Battery Using Mask R-CNN

Yih-Lon Lin, Yu-Min Chiang, Chia-Ming Liu, and Sih-Wei Huang

Abstract Recently the growing sensitivity of various governments toward a cleaner environment has increased the demand for electric vehicles (EVs). The battery is one of the most vital components in an EV. With the rapid development of EVs, the applications of lithium-ion battery (LIB) have become more and more extensive. Among the LIBs, the pouch type lithium-ion battery offers a simple, flexible, light weight, and robust solution to battery design, therefore it is considered to be the most promising technology for power battery. The pouch type LIB cell manufacturing processes include lithium battery cell assembly, electrolyte filling, formation, and aging, etc. The purpose of the formation is to form a stable SEI film on the electrode surface by charging. In the formation process, if the electrode tabs deviates from the accepted range, that will cause the failed charging. Therefore, a machine vision system should be built to automatically detect the electrode tab deflection.

Recent developments in the field of deep learning have inspired a new interest in using neural network for general image classification tasks. In this paper, we adopt an instance segmentation algorithm called Mask R-CNN to detect and segment electrode tabs. Using the generated bounding boxes of the Mask R-CNN together with the Canny edge detection method, we can decide whether the electrode tab has deflected. As there was no existing dataset, we built a new dataset containing 398 pouch LIBs images. Among them, 200 samples were used as the training dataset, and the other 98 images were used as the validating dataset. Experimental results show 100% accuracy in electrode tab detection for both training and validating dataset. Another 100 pouch LIBs images were captured and used as test dataset. For the test dataset, the proposed algorithm has 100% accuracy in electrode deflection detection

Y.-L. Lin
National Yunlin University of Science and Technology, Yunlin, Taiwan

Y.-M. Chiang (✉)
National Formosa University, Yunlin, Taiwan
e-mail: ymchiang@nfu.edu.tw

C.-M. Liu · S.-W. Huang
I-Shou University, Kaohsiung, Taiwan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_14

and 95% accuracy in polarity. It verifies the effectiveness of the proposed approach in electrode tab deflection detection for pouch LIB.

Keywords Deflection detection · Mask R-CNN · Pouch lithium-ion battery · Eelectrode tab · Deep learning

1 Introduction

Driven by the ecological effects of CO2-emissions, the electrical drive technologies have become as the most popular transportation applications [1]. With an increasing demand of electric vehicles (EVs), the applications of lithium-ion battery (LIB) have become more and more widespread.

Schröder et al. [2] gave a detailed overview on the pros and cons of different shapes of LIBs, including cylindrical and prismatic geometry, whereas the prismatic shape can be further divided into hard-case cell and pouch cell. The pouch cell manufacturing process of cutting increases the risk of manufacturing errors. In terms of safety, the housing stability of the pouch cell is considerably lower than the cylindrical and the hard-case cell. However, compared to the cylindrical and prismatic hard-case cell, the pouch cell has distinct advantages in terms of energy and packing density, customized design shape, as well as manufacturing costs. It leads to a variety of applications in electric vehicles and the consumer electronic products.

The pouch lithium-ion battery cell manufacturing process includes cell assembly and cell activation, as shown in Fig. 1. In the cell assembly process, the cathode (positive) material and the anode (negative) material are mixed and coated on copper foil and aluminum foil, and then the foils are made into rolls. The thickness of the metal foil is controlled by rolling to be uniform. After the metal rolls are dried, they are cut into desired shape to form electrodes. Subsequently to the cutting process, the compound for the pouch cell can either be created by laminating single sheets of electrodes and separator onto each other or by creating a z-folded separator structure [3]. The tab welding electrically connects the assembled electrodes. The housing for the pouch cell is a foil made of a deep-drawn aluminum-polymer-compound, a so called aluminum pouch foil. The laminated compound is simply inserted onto the pouch foil, and then sealed at three edges.

The main processes of cell activation comprise of electrolyte filling, formation, and aging [2]. The formation is the initial charging and discharging process for the battery cell. During the formation process, gas forms and builds up within the pouch foil housing. The pouch cell relies on an additional attached pouch bag, designed to absorb the gas formation [2]. The gas filled pouch bag is vacuumized subsequently then the pouch cell is sealed. The aging process is the final process step of the cell manufacturing. It is designed to measure the performance and the properties of the battery cells. After final testing and visual inspection, the battery cells are ultimately graded according to their performance and properties. Finally the graded pouch cells are packed and shipped to the customers.

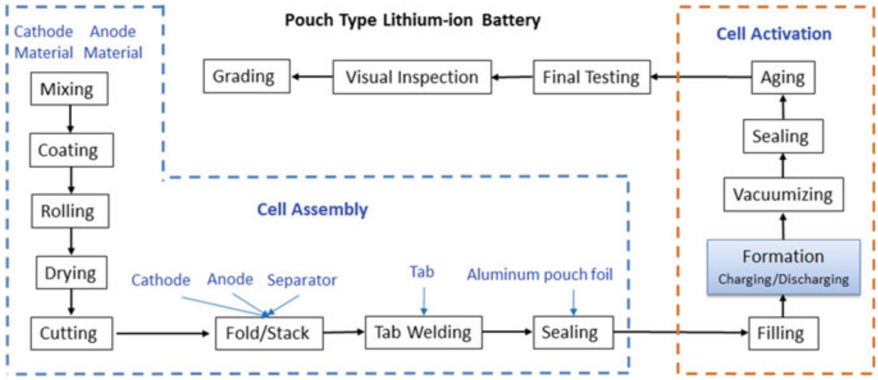


Fig. 1 The manufacturing processes of pouch type lithium-ion battery

Due to the growing demand, an automated production line for pouch lithium-ion cell is necessary for LIB manufacturers as it may achieve the goals of improving quality, increasing yield and reducing cost. The inspection done by manual operation or an automated system helps to identify defects or faults, which could result in malfunction of the LIB. Once the automated systems have replaced the human operators, they are called the automated optical inspection (AOI) systems and are often part of the automated production lines.

In pouch cell activation process, the purpose of the formation is to form a stable SEI film on the electrode surface by charging. In the formation process, if the electrode tabs deviates from the accepted range, that will cause the failed charging. Therefore, a machine vision system should be built to automatically detect the electrode tab deflection.

Recent developments in the field of deep learning [4] have inspired a new interest in using neural network for general image recognition tasks, such as face recognition [5–7], medical image application [8–10], and defect detection in industry application [11–13]. For the object detection applications, it can be found deep convolutional neural networks (DCNNs) are on the cutting edge. However, not all approaches are designed for real-time applications. To fast detect an object, some specific DCNNs were proposed, such as Fast R-CNN [14], Faster R-CNN [15], and YOLO [16]. Based on the Faster R-CNN model, a novel object detection algorithm called Mask Region-CNN (R-CNN) [17] has been proposed, which can not only classify and localize the object, but also can segment the object at the same time.

In the manufacturing processes of pouch LIB, visual inspection is responsible for detecting defects to ensure quality for all finished products, and in general it needs a large amount of labor cost. Lin et al. [18] proposed a defect detection and segmentation method based on Mask R-CNN to inspect pouch LIB defects. Inspired by Lin et al. [18], in this paper, we will develop a method to detect and segment electrode tabs by using Mask R-CNN.

This article is organized as follows. Section 2 discusses in detail the proposed deflection detection for electrode tabs based on mask R-CNN and Canny edge detection[19], followed by the implementation and experimental results as well as the conclusion.

2 Electrode Tab Deflection Detection

2.1 Consideration for Vision System

Compared to the cylindrical and the hard-case cell, the number of electrodes to be tailored for manufacturing a pouch cell is much higher. As the electrode tabs are sheet type, they are easy to deflect in welding and sealing processes of cell assembly. Subsequently, in the formation process, if the electrode tabs deviates from the accepted value (in our case, the value is ±3 mm), that will cause the failed charging. Figure 2 shows twelve pouch cells placed in a set of fixtures for formation. The upper tabs are negative (anode) electrodes, and the lower tabs are positive (cathode) electrodes. The camera is located above the middle of the set of fixtures. Since the electrode tab is at least 30 mm high, and due to the linear perspective created when taking image, it can be observed that the size of each electrode tab is different. To detect the deflection of electrode tab, this paper suggested capture only one pouch cell to avoid the deformation of the electrode tab. The captured image is shown in Fig. 3.

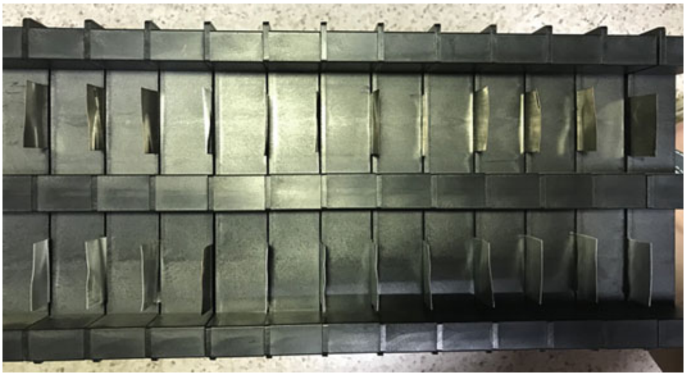


Fig. 2 Captured image of 12 pouch cells in a set of fixtures

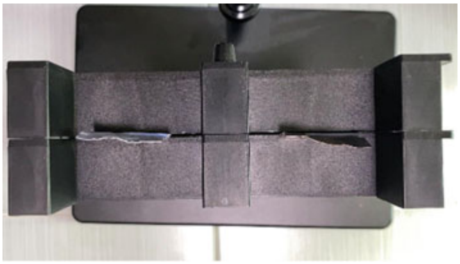


Fig. 3 Captured image of only one pouch cell

2.2 Mask R-CNN Based Electrode Tab Deflection Detection

As the electrode tabs of pouch cell is easy to reflect light, and the deflection are with high degree of variability, this study try to adopt a mask inspection network method to detect electrode tab deflection. Since there is no publicly available dataset, we collected images such as shown in Fig. 3 by a machine vision system.

This research proposed an approach to detect electrode tab deflection by using an instance segmentation deep learning model called Mask R-CNN, which can detect and segment electrode tabs. Followed by Canny edge detection method, the proposed method therefore can detect the electrode polarity and the deflection simultaneously. The following introduce the Mask R-CNN method briefly.

The architecture of Mask R-CNN can be seen in Fig. 4. Mask R-CNN consists of several modules. The first is a standard CNN that serves as a feature extractor, called backbone. In principle, this backbone network can be any network that extracts image features. In our model, ResNet-101 is used as the backbone. In addition, to realize multi-scale detection, a feature pyramid network (FPN) is also used in the backbone network.

Mask R-CNN relies on region proposals which are generated via a region proposal network (RPN). The RPN is a lightweight neural network that scans the image in a sliding window and finds candidate bounding boxes of objects. The regions that the

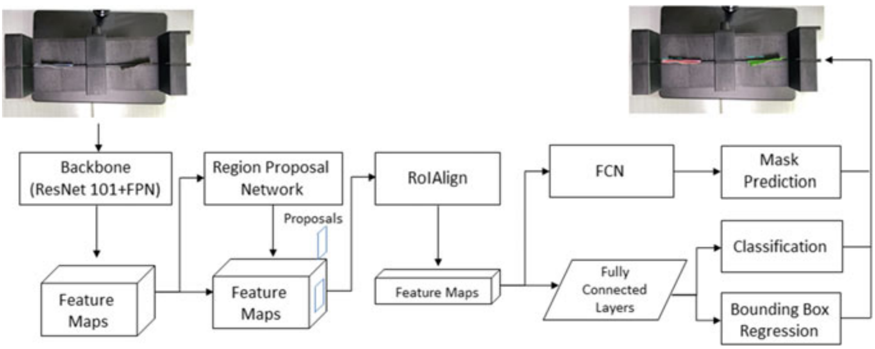


Fig. 4 Mask R-CNN framework for electrode tab detection of pouch cell

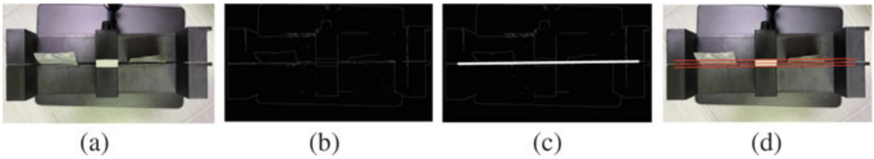


Fig. 5 Canny edge detection process: **(a)** input image, **(b)** result of Canny edge detection, **(c)** drawing the center line, **(d)** drawing the control line

RPN scans over are called anchors. Anchors that overlap too much will be screened according to non-max suppression (NMS). The threshold value for bounding-box NMS is usually set to 0.5.

Mask R-CNN replaces the somewhat imprecise RoI (regions of interest) pooling operation used in Faster R-CNN with an operation called RoIAlign. RoIAlign eliminates the process of integral and preserves decimals, thus it is more accurate for detection and segmentation. Mask R-CNN adds a separate fully convolutional network (FCN) head to produce the desired instance segmentations, as illustrated in Fig. 4. The mask network head predicts the mask independently from the network head predicting the class and bonding box. This leads to the use of a multi-task loss function combining the loss values of classification, bonding box regression, and segmentation.

After the electrode tab segmentation process by applying Mask R-CNN, this study further adopted the Canny edge detection operator to detect the base line for deflection detection. Once the base line of the original image is detected, we draw two control lines for detecting the deflection, as shown in Fig. 5. The location of the control lines are computed by the information got from prior camera calibration. If the bonding boxes of the electrode tabs out of the control line, it will be classified to NG. Figure 5 illustrates the deflection detection processes using Canny edge detector.

3 Experiments

Since there is no publicly available dataset for deflection detection for pouch LIB, we collected images using a machine vision system which is installed parallel to the electrode tabs and the field of view (FOV) is only one pouch cell. The original image size is 1920 × 1080 (px) and is resized to 1024 × 1024 (px) for Mask R-CNN model training. Totally we collected 298 sample images. Among them, 200 samples were used to train the Mask R-CNN model and the other 98 samples were used as validation dataset.

After image collection, the research utilized the Lableme for annotation [20]. In our application, the objects include anode electrode tab (red) and cathode electrode tab (green), as shown in Fig. 6. The vertex coordinates for one image are saved in a JSON file. The labeling process is a manual operation and may not be very precise. However, it will not affect the evaluation of the model overall. The pixels inside

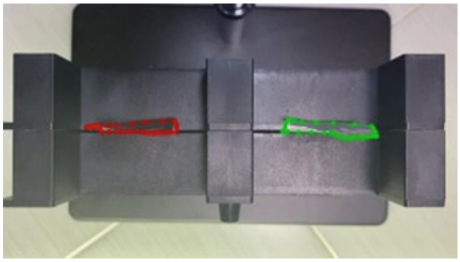


Fig. 6 Labeling for electrode tab

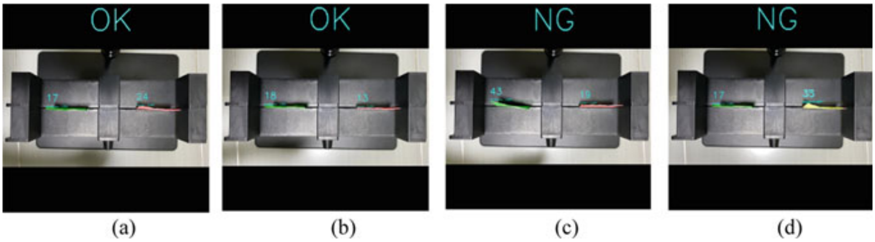


Fig. 7 Some electrode tab deflection detection results of test dataset

the bounding polygons corresponding to the two types of electrode tabs are given a value 1 and 2, and the rest of the pixels are regarded as background with a value 0.

The proposed approach was implemented using python language based on the open-source libraries Keras, Tensorflow, and OpenCV. We used a ResNet-101 feature pyramid network model as a backbone. In total we trained for 320 epochs using stochastic gradient descent (SGD) and a batch size of 2 on a single NVIDIA 2080 Ti GPU. For the training and validation dataset, the Mask R-CNN successfully detects electrode tabs with 100% accuracy. To further verify the effectiveness of the proposed deflection detection algorithm, another 100 pouch cell images were captured under different lighting condition as test dataset. The test dataset consists of 37 deflection samples. Some detection results of test dataset are shown in Fig. 7. Though the classification result indicates 5% error rate in polarity (Fig. 7d), the pouch cells with deflected electrode tabs were all recognized.

4 Conclusion

Lithium-ion battery system plays an important role in the development of electrical vehicles. The pouch cell has distinct advantages in terms of energy and packing density, customized design shape, as well as manufacturing costs as comparing to the cylindrical and prismatic hard-case cell. It leads to a variety of applications in electric vehicles and the consumer electronic products. In pouch cell activation process, the purpose of the formation is to form a stable SEI film on the electrode surface by charging. In the formation process, if the electrode tabs deviates from the accepted range, that will cause the failed charging. Therefore, a machine vision system should be built to automatically detect the electrode tab deflection. In this paper, we studied the potential of using the instance segmentation techniques, Mask R-CNN, to the electrode tab deflection detection for the formation process of pouch lithium-ion cell manufacturing. By implementing a Mask R-CNN model and using Canny edge detection, it shows a remarkable performance in electrode tab deflection detection.

Acknowledgements This research was supported by the Ministry of Science and Technology, Taiwan, under grant no. MOST 108-2221-E-507-003- and MOST 109-2221-E-507-008-.

References

1. Hannan, M.A., Hoque, M.M., Mohamed, A., Ayob, A.: Review of energy storage systems for electric vehicle applications: issues and challenges. Renew. Sustain. Energy Rev. **69**, 771–789 (2017)
2. Schröder, R., Aydemir, M., Seliger, G.: Comparatively assessing different shapes of lithium-ion battery cells. Proc. Manuf. **8**, 104–111 (2017)
3. Reinhart, G. et al.: Research and demonstration center for the production of large-area lithium-ion cells. In: WGP, Berlin, pp. 3–12 (2011).
4. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
5. Parkhi, O. M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference (BMVC), vol. 1, pp. 1–12 (2015)
6. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multi-task cascaded convolutional networks. Comput. Vis. Pattern Recogn. arXiv:1604.02878 (2016)
7. Wang, M., Deng, W.: Deep face recognition: A survey. arXiv:1804.06655 (2018)
8. Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J.N.L., Isgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans. Med. Imaging **35**(5), 1252 (2016)
9. Wolterink, J. M., Kamnitsas, K., Ledig, C., Išgum, I.: Generative adversarial networks and adversarial methods in biomedical image analysis. Comput. Vis. Pattern Recogn. arXiv:1810.10352 (2018)
10. Kazeminia, S., Baur, C., Kuijper, A., Ginneken, B. V., Navab, N., Albarqouni, S., Mukhopadhyay, A.: GANs for medical image analysis. Comput. Vis. Pattern Recogn. arXiv:1809.06222 (2018)
11. Yi, L., Li, G., Jiang, M.: An end-to-end steel strip surface defects recognition system based on convolutional neural networks. Steel Res. Int. **88**(2), 176–187 (2017)
12. Tsai, D.M., Chou, Y.H.: Fast and precise positioning in PCBs using deep neural network regression. IEEE Trans. Instrum. Meas. **69**(7), 4692–4701 (2020). https://doi.org/10.1109/TIM.2019.2957866
13. Jing, J.F., Ma, H., Zhang, H.H.: Automatic fabric defect detection using a deep convolutional neural network. Color. Technol. **135**(3), 213–223 (2019)
14. Girshick, R.: Fast R-CNN, In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
15. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. NIPS (2015)

16. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2017)
17. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN, In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988 (2017)
18. Lin, Y.L., Chiang, Y.M, Huang, S.W., Liu, C.M.: Defect detection and segmentation for pouch Lithium-ion battery using mask R-CNN, In: The 17th International Conference on Automation Technology (Automation 2020), Hualien, Taiwan.12–14 November 2020
19. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
20. Russell, B., Torralba, A., Freeman, W. T.: Labelme, the open annotation tool MIT. Computer Science and Artificial Intelligence Laboratory [Online]. Available: http://labelme.csail.mit.ed

Comparison of Process Chains of Additive and Conventional Manufacturing

Nikolas Zimmermann, Joachim Lentes, Sascha Schaper, and Andreas Werner

Abstract Numerous companies are thinking about using additive manufacturing in their production processes to leverage expected potentials such as higher flexibility, reduced costs around tools, jigs, and spare parts as well as a reduction of costs and times for manufacturing of complex goods. A substitution of an existing subtractive production process is what comes to their mind most of the time. They hope for a better material efficiency and an easier and cheaper way to produce small batches of products. The factual circumstances, however, are often much more complex than some companies, especially small and medium-sized enterprises (SME), initially assume. To give companies a clearer idea of the implications of using additive manufacturing, a rough filter model and a comparison of process chains of additive and conventional manufacturing is presented in this paper. The analysis consists of both the digital as well as the physical process chain and thereby provides a holistic picture of changes that are implied by the usage of additive manufacturing. In order to provide a clear reference for manufacturing companies, the report also discusses facts that have been discovered in cooperation with an SME during the production of real customer parts.

Keywords Additive manufacturing · Process chain · Efficiency

1 Introduction

Small and medium-sized enterprises (SME) form the core of the manufacturing industry in Germany and are often established hidden champions. Nevertheless, they face a lot of different challenges to retain their place in the market. The former digitalization of the production, limited resources, and a shorter time-to-market are only

N. Zimmermann (✉) · J. Lentes · S. Schaper
Fraunhofer Institute for Industrial Engineering IAO, Nobelstrasse 12, 70569 Stuttgart, Germany
e-mail: nikolas.zimmermann@iao.fraunhofer.de

A. Werner
Institute of Human Factors and Technology Management, University of Stuttgart IAT, Nobelstrasse 12, 70569 Stuttgart, Germany

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_15

a few of them. Additive manufacturing (AM), also known as 3D-printing, offers the possibility to efficiently address these challenges when used correctly. However, the circumstances to implement AM in the own company are much more complex than initially assumed by most companies. In order to give them a better understanding of the necessary changes, this paper presents different efficiency aspects of AM and compares its process chain to a generalized one for conventional techniques.

1.1 Previous Research

In the past years, the number of applications for AM has significantly increased and thereby the technology is established in many different industries. This is why related research activities are still growing, and various fields are analysed by scientists all over the world. Current studies mainly focus on the technology for additive manufacturing itself and the materials used in the manufacturing process. Publications regarding profitability and sustainability are outnumbered [1].

The researchers that compare AM to conventional manufacturing analysed mainly three key attributes: Their general outcome is, that AM is better suited for tasks with high customizability and complexity whether traditional methods have advantages in production volume [2]. The reason for this is the high design freedom for AM parts. Thus, the engineers can focus more on the intended use of the products than the manufacturing restrictions [3]. However, the limits and coincidently the strengths of subtractive manufacturing are precision, surface quality and reliability. Further, depending on the AM technology and the use case of the part, additional post-processing steps could be necessary to achieve design requirements [4, 5].

Comparing the manufacturing methods on a more detailed level, the production costs are often the focus of research. Various studies compare different AM technologies with injection molding regarding the fabrication of plastic parts. Hopkinson & Dickens are among the first who analyzed the major sources of cost for AM [6]. They divided them into machine, labor, and material costs. Subsequent studies mostly took their work as a basis. For example, Ruffo et al. developed a comprehensive cost model including overhead costs. They found out that particularly the machine costs, which depend on the production time, are one of the main cost drivers [7]. Achillas et al. compare PolyJet, selective laser sintering (SLS), stereolithography (SLA), fused deposition modelling (FDM) and injection molding regarding lead time and total production cost. Up to 500 parts, injection molding is the most time-consuming technology due to the mould, which needs to be designed and built. The lead time of AM technologies is dominated by their process time caused by their relatively low building speed. According to their study, SLS is overall the most time- and cost-effective AM technology [8].

When comparing additive and conventional manufacturing, not only the costs but also the environmental impact of the technologies needs to be investigated. There are some applications where AM is confirmed to be more environmentally friendly, but the assessment is very complex and depends heavily on the process used [9]. However, this assumption is deceptive because more factors should be taken into consideration. In most studies, a life-cycle assessment (LCA) is carried out to determine parameters such as energy consumption, material waste and emissions. Faludi et al. compare different AM technologies with CNC milling and point out that none of the methods is categorically more sustainable. It always depends on the part design and which machine is used. They state that among the evaluated AM methods, FDM is the most environmentally friendly [10]. Paris et al. use a LCA to capture the environmental impact of Electron Beam Melting (EBM) and milling. They conclude that EBM is more sustainable for parts with complex geometries and shapes which would require a lot of material removal with subtractive technologies. In addition, they ascertain that the energy consumption of both methods is nearly identical [11]. Regarding additional emissions, Brent et al. investigated ultrafine particle output of desktop 3D-printers and characterized them as "high emitters". He suggests caution when operating them in unvented rooms [12]. The German Federal Environment Agency confirms this problem with their study and further points out the high energy consumption of AM machines. On the other hand, they see potential regarding resource efficiency [13].

1.2 Objective and Approach

With regard to previous research activities, it is made clear that most studies focus on a specific topic regarding AM, which often consider many details and are not suitable for a direct transfer into common practice. To make AM more accessible to SME, a more holistic approach needs to be developed. Thus, the objective of this paper is to enable SMEs to efficiently use AM for engineer-to-order tasks. This is addressed by using a systematic efficiency approach to define a filter model, which contains three decision guidance's steps. This proceeding is the basis to evaluate if AM is the best choice for a given customer order and consequently, how efficiently it can be handled with AM. In the next step, the process chains of additive and conventional manufacturing are compared to each other. With this analysis, differences in the physical and digital process steps are identified. This section is complemented by discoveries from the practice during the production of real customer parts. Hereby, SMEs should get a good overview over the changes that are necessary to use AM in their own company. The contribution ends with a conclusion and outlook on further work.

2 Efficiency Approach for Additive Manufacturing

Many SMEs want to use AM in their production processes and thus exploit the numerous potentials of this new technology. In particular, the high degree of design freedom and the possibility of function integration in relation to the product are advantages that can be realized relatively easily [14]. Nevertheless, a suitable product, taking manufacturability into account, is the basis for an efficient production.

Companies often underestimate the necessary changes in their overall production and the duration until AM can be used efficiently. To determine whether an engineer-to-order task is really a suitable use case for AM and which potentials can ultimately be achieved, a filter model is introduced (Fig. 1). Before the start of production or in the planning phase, the customer order must pass three evaluation steps and thus has to show the fulfillment of three requirements: feasibility, economic viability and sustainability. Each of these requirements has several different attributes that must be considered when using AM. If the production order passes all three filters, it can be considered as an efficient use case for AM. The better the three requirements are met, the more potential can be exploited and the fewer risks come into play. In the end, it should be clear in which steps AM has been beneficial and in which steps there is still need for action to fully exploit the strengths and opportunities.

Going one step further, an evaluation mechanism needs to be developed for all three filters to systematically determine the degree of efficiency. Since this undertaking is beyond the scope of this paper, only the process chain will be analyzed in more detail as the main part of this study. Nevertheless, a brief overview of the three filters will be given in the next sections.

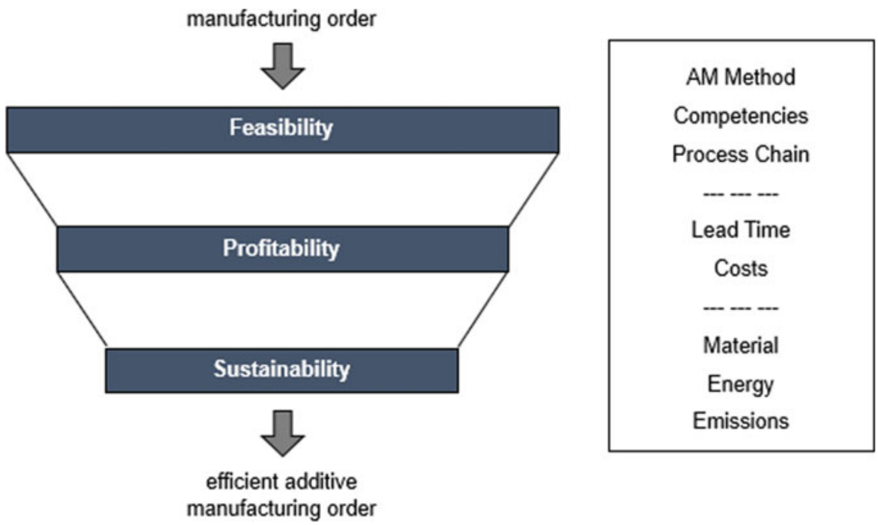


Fig. 1 Filter model for order evaluation

2.1 Feasibility

In the first step, the customer order is checked for feasibility. As a basis for all subsequent requirements, it must be determined with which AM method the product can be manufactured and whether this method is established in the company. There are several manufacturing parameters for the various AM processes, such as positioning, orientation, slicing and hatching. When starting from scratch, it may take some time before parts can be built reliably and with sufficient quality. If AM itself cannot be fully implemented, another manufacturing method should be chosen. In addition, the skills of the employees must match those required for AM. When substituting subtractive manufacturing with additive manufacturing, it is necessary to match the skills of the employees. Due to the differences in technology, an employee who previously operated a Computerized Numerical Control (CNC) milling machine cannot operate a 3D printer in a simplified manner. After evaluating the required AM method and all necessary competencies, the process chain must be considered. A detailed analysis of the process steps is performed in Chapter 3 of this paper. According to these three areas, a manufacturing job is feasible if the appropriate AM method, the right competencies, and a continuous process chain are available. With this approach, a preliminary check is carried out, which does not yet evaluate the profitability of the order but checks whether it is capable of being manufactured.

2.2 Profitability

The profitability of the application of AM is the next filter to evaluate if an order can be handled with AM. Nowadays, a shorter time-to-market and a reduced product lifecycle are challenges, which companies face. The consequence is the need for a short lead time and a cost-effective production. Under certain circumstances, AM can help to achieve these goals. Therefore, it must be assessed if an AM process is less time-consuming than a conventional one. As already mentioned, the build time has the largest share of the entire lead time. Several studies show how it can be estimated in advance, for example by approximating the geometry with voxels [15]. Based on the lead time, a comprehensive cost model is needed to further determine the economic aspect of the order. An overview of different models and how they evolved over time is given by Costabile et al. [1].

2.3 Sustainability

After a customer order is proved as efficient regarding the economic aspect, nowadays there is also a need regarding its sustainability. Since manufacturing processes use natural resources to produce goods, it should be compared how much they consume and in which way they influence the environment and the society. The material is one of the main aspects, which has impact in this regard. While the choice of material often comes hand in hand with the choice of the manufacturing method, it is important to keep aspects like waste and recyclability in mind. Especially the non-degradability of plastics is a big issue. Researchers already propose different ways to recycle polymers and use them as a feedstock for 3D printers. The implementation of AM in a circular economy and the use of local material supply chains offer big potentials, when considered early enough [16, 17]. The second aspect which comes into account, is the energy consumption of AM. Depending on the AM method, the machines still consume a lot of energy, as already pointed out in chapter 1.1. Therefore, the energy efficiency is a huge adjustment possibility to make AM more sustainable [13]. Consequently, emissions also play an important role. On one hand the carbon dioxid (CO_2) emissions, which depend on the previously mentioned energy consumption, on the other hand, ultrafine particles are also an issue.

3 Analysis of Process Chains

The idea behind the analysis of the process chains for conventional and additive manufacturing was derived through the practical experience that especially SME are underestimating how fundamentally some steps differ from each other when comparing the process chains. Obviously, the process chains must be kept relatively general, since the details and scope of the individual steps can vary greatly depending on the manufacturing process selected.

3.1 Process Chains

The process chains of subtractive and additive manufacturing are depicted in Figs. 2 and 3. The processes are divided into five steps. Three of them, namely the Pre-, In- and Post-Process, form the core and directly refer to the manufacturing method. Even though the representation of the subtractive process chain in this manner is not common, it is chosen to enable a better comparison of the different activities in chapter 3.2.

The development comprises steps to design a suitable product. This phase is generally further divided and a procedure model is used to make the development process manageable. Even if it is an important step of the product creation, it is not an element of the actual production. Therefore, it is only mentioned briefly as a basis for all following steps. The result of the development is a document, which carries all the necessary information for the manufacturing of the part.

The Pre-Process is the first and the most crucial step of the production. The decisions that are made in this step heavily influence the quality of the product. In addition, a good preparation enables a fluent production without unforeseen interruptions and

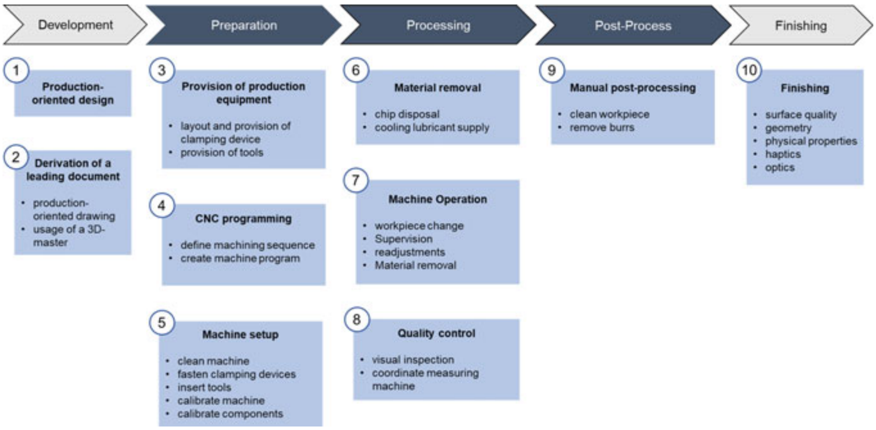


Fig. 2 Process steps for subtractive manufacturing

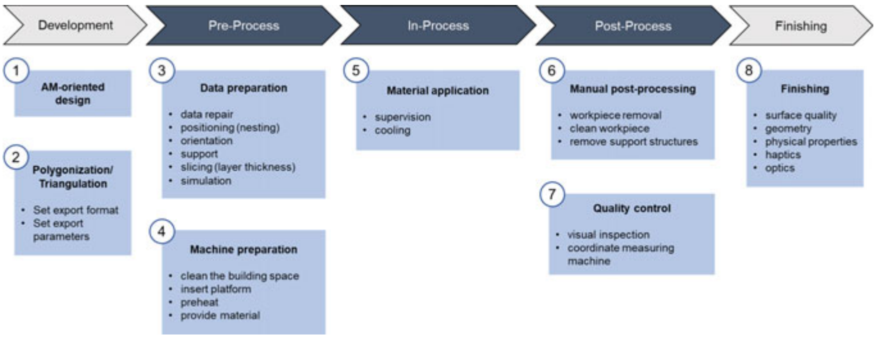


Fig. 3 Process steps for additive manufacturing

increasing costs. The activities involved split up into physical steps like machine preparation and digital steps such as data preparation or CNC programming.

The actual production of the workpiece takes places in the next step where the product is in-process. This part is heavily influenced by the chosen manufacturing method, accordingly if material is removed or applied. However, in both cases supervision is necessary to ensure that mistakes are discovered as fast as possible.

After the processing, the post-process is carried out. Several manual activities are necessary to bring the product in its final form. For this purpose, the workpiece is removed from the machine and cleaned. Further steps sometimes involve heat treatments or tempering to achieve certain material properties. In addition, a quality control needs to be carried out to determine if the product meets the requirements. A visual inspection is able to detect most of the surface faults. For a more detailed evaluation, other techniques are more appropriate.

In the last phase, optional finishing steps could be necessary. These are generally conditioned by part specifications, which are not readily achievable with the selected manufacturing technology. For both process chains this might involve a second manufacturing process with a different method. For that reason, the finishing is only mentioned briefly but not covered in detail.

3.2 Comparison

After analyzing the two different process chains, it is noticeable that they are quite different. In the following, nine main differences are identified and examined in more detail (Fig. 4).

(a) Major changes in the product and its development are necessary to leverage the potential of AM. Parts for conventional methods are restricted by the manufacturability. Parts for AM often have more design freedom and the function of the product decides its design. However, design guidelines are only in the beginning and not as detailed as for conventional parts. AM might change the way how products are designed by enabling a more agile development and replacing the common stage-gate approach.

(b) AM is inherently digital. The product data is transferred directly from a 3D-CAD program into the printer software. There is no need for a 2D drawing, which mostly still is the leading document for conventional manufacturing and enables machine operators to manufacture the product. With AM nearly all decisions regarding the manufacturing process are made in the pre-Processing phase.

(c) AM does not require tools. Conventional manufacturing methods like milling need multiple tools for the material removal. In addition, clamping devices are needed to fasten the workpiece inside the machine. For other shaping manufacturing technologies like injection moulding a mould must be designed and built before the actual production of parts can start. This requires a lot of time and

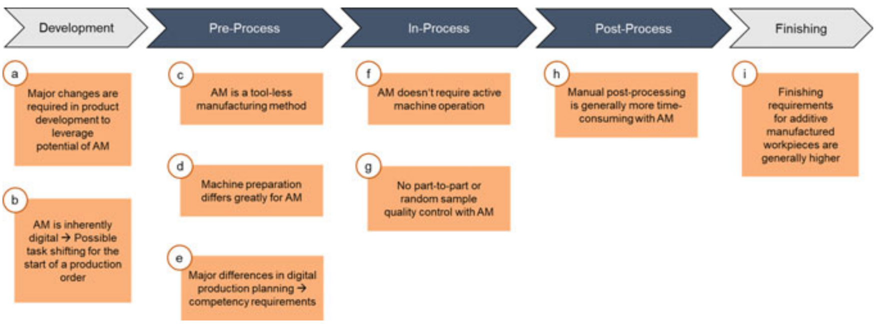


Fig. 4 Overview of main differences between AM and conventional manufacturing

increases costs, being the main reason why AM can be more cost-effective for small batch sizes.

(d) The entire procedure for preparing an AM system differs fundamentally from conventional processes. The main task with AM is the material supply, which can be filament, powder or resin supply, depending on the process. Many activities that are necessary in conventional processes are eliminated without replacement. However, depending on the process, calibration processes are necessary at certain time intervals.

(e) The production planning for AM is digital and often more complex in comparison to conventional manufacturing. That is a result of the various setting parameters during the data preparation. To understand the different effects of each parameter on the final product, deep knowledge and experience is required. In addition, standards and guidelines are still rare.

(f) AM does not require active machine operation. During the printing process no action is needed unless the material application fails. The printer can run relatively independently. On the contrary, CNC milling machines require supervision and usually regular adjustments. Manual part change is still standard for small series.

(g) AM quality control is limited and not as advanced as for conventional manufacturing. Methods like Acceptable Quality Limit sampling control or Statistical Process Control are common parts of the quality management systems of SME. However, these tools do not have established for AM yet. In addition, the process chain must be designed in such a way that the process quality and the resulting product quality are consistent and reproducible. Due to the experience of the SMEs, the conventional workflow is robust and usually not very error-prone. For AM, this goal is much more complex to reach and will require additional quality aspects like trained staff, comprehensive documentation, and feasible infrastructure [18].

(h) Post-Processing is generally more time-consuming with AM. One reason is that support structures must be removed manually, either mechanically by hand or chemically by dissolving them. In addition, for methods with power bed the cleaning is quite costly. In contrast, post-processing steps for milled parts like washing and deburring are usually automated and therefore do not require additional work effort.

(i) Finishing requirements for AM are generally higher. Compared to conventional manufacturing, the achievable surface quality is quite rough, and tolerances are harder to maintain depending on the process. Therefore, functional surfaces have to be post-processed accordingly. Examples comprise bearing seats or running surfaces. With conventional manufacturing, the required tolerances can mostly be achieved right away without any external finishing steps.

3.3 Exemplary Discrepancies in Process Chains of a Manufacturing SME

The following two examples occurred in the execution of production orders of an SME. The point of providing these two examples here is the apparent simplicity of AM, which is easily faltered by unforeseen deviations.

(a) The first example shows an irregularity in the digital process chain. During the execution of a production order from an automotive OEM the requested file was sent to the SME as a STEP (Standard for the Exchange of Product Model Data) file – the STEP file was directly exported from the Product Lifecycle Management (PLM) system of the OEM, resulting in an error in the printing program. The geometry could not be recognized correctly. The error could be corrected by first opening the file in a CAD authoring system and then exporting it as a STEP file. This is an example for unforeseen deviations, which delayed the production start of the part and shows that there are still learning curves regarding the usage of AM.

(b) The second example illustrates the task shifts described in chapter 3.2. b). During the production process of the fitting shown in Fig. 5, the responsibility for component positioning was initially with the machine operator on the production floor. The fitting was positioned in the installation space as shown on the left, which meant that the holes were not round due to the layer-by-layer production. The necessary competence on the part of the machine operator was therefore lacking. In the case of AM, the positioning of the component and the start of production can easily be transferred to indirect areas, since AM is an inherently digital manufacturing technology. The illustration on the right shows the correctly positioned part, the assembly could be disassembled directly in the printing software by using proprietary data formats.

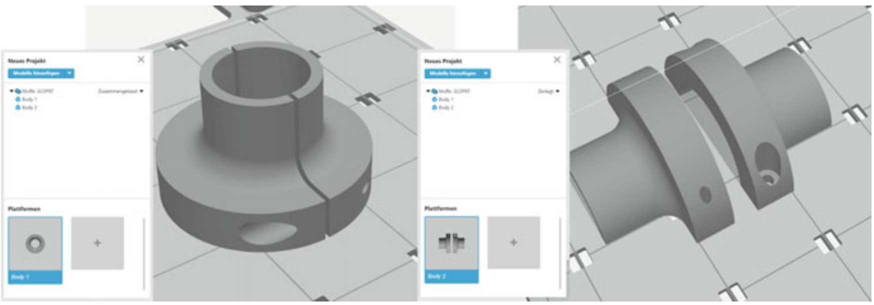


Fig. 5 Positioning of workpieces

4 Conclusion

The comparison between the process chains gives a first overview of major differences between AM and conventional production processes. With the filter model, SME can initiate an assessment of whether production orders should be manufactured additively or conventionally. The findings during the collaboration with an SME during the project still revealed a few discrepancies which could not been foreseen during the analysis phase. In the course of the project, the aspect of sustainability will be further examined in detail, as well as the influences of the different factors on each other.

Acknowledgements The content of this paper is part of the work in the project "Ultraeffizienz Fabrik" - Establishment of a hybrid center for ultra-efficient factories, which has received funding from the Ministry of the Environment, Climate Protection and the Energy Sector Baden-Württemberg.

References

1. Costabile, G., Fera, M., Fruggiero, F., Lambiase, A., Pham, D.: Cost models of additive manufacturing: A literature review. Int. J. Ind. Eng. Comput. **8**, 263–282 (2017)
2. Pareira, T., Kennedy, J., Potgieter, J.: A comparison of traditional manufacturing vs additive manufacturing, the best method fot the job. 14th Global Congress on Manufacturing and Management. Procedia Manufacturing **30**, 11–18 (2019).
3. Kumke, M.: Methodisches Konstruieren von additiv gefertigten Bauteilen, Dissertation der TU Braunschweig, Springer Fachmedienverlag, (2019).
4. Häfele, T., Schneberger, J.-H., Kaspar, J., Vielhaber, M., Griebsch, J.: Hybrid additve manufacturing—process chain correlations and impacts. Proc. CIRP **84**, 328–334 (2019)
5. Kaspar, J., Stoffels, P., Schneberger, J.-H., Vielhaber, M.: Integrated product, production and material definition for conventional versus generative manufacturing technologies. Proc. CIRP **70**, 180–185 (2018)
6. Hopkinson, N., Dickens, P.: Analysis of rapid manufacturing - using layer manufacturing processes for production. Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. **217**(1), 31–39 (2003)
7. Ruffo, M., Tuck, C., Hague, R.: Cost estimation for rapid manufacturing - laser sintering production for low to medium volumes. Proc. Inst. Mech. Eng. Part B J. Eng. Manuf. **220**(9), 1417–1427 (2006)
8. Achillas, C., Tzetzis, D., Raimondo, M.O.: Alternative production strategies based on the comparison of additive and traditional manufacturing technologies. Int. J. Prod. Res. **55**(12), 3497–3509 (2017)
9. Kellens, K., Baumers, M., Gutowski, T.,Flanagan, W. Lifset, R., Duflou, J.: Environmental dimensions of additive manufacturing. J. Ind. Ecol. **21**(1) (2017)
10. Faludi, J., Bayley, C., Bhogal, S., Iribarne, M.: Comparing environmental impacts of additive manufacturing vs traditional machining via life-cycle assessment. Rapid Proto. J. **21**, 14–33 (2015)
11. Paris, H., Mokhtarian, H., Coatanea, E., Museau, M., Ituarte, I.: Comparative environmental impacts of additive and subtractive manufacturing technologies. CIRP Annals. Manuf. Technol **65**, 29–32 (2016)
12. Brent, S., Azimi, P., El Orch, Z., Ramos, T.: Ultrafine particle emissions from desktop 3D printers. Atmos. Environ. **79**, 334–339 (2013)

13. Keppner, B., Kahlenborn, W., Richter, S., Jetzke, T., Lessmann, A., Bovenschulte, M.: Die Zukunft im Blick: 3D-Druck/Trendbericht zur Abschätzung der Umweltwirkungen. Umweltbundesamt 2019
14. Zimmermann, N., Lentes, J., Werner, A.: Analysis of requirements, potentials and risks caused by using additive manufacturing. Proc. Manuf. **39**, 474–483 (2019)
15. Baumers, M., Tuck, C., Wildman, R., Ashcroft, I., Rosamond, E., Hague, R.: Combined buildtime, energy consumption and cost estimation for direct metal laser sintering. In: 23rd Annual International Solid Freeform Fabrication Symposium—An Additive Manufacturing Conference, SFF, pp. 932–944 (2012)
16. Sanchez, F.A.C., Boudaoud, H., Hoppe, S., Camargo, M.: Polymer recycling in an open-source additive manufacturing context: Mechanical issues. Addit. Manuf. **17**, 87–105 (2017)
17. Sanchez, F.A.C., Boudaoud, H., Camargo, M., Pearce, J.M.: Plastic recycling in additive manufacturing: A systematic literature review and opportunities for the circular economy. J. Clean. Prod. **264**, 121602 (2020)
18. DIN SPEC 17071:2019–12, Additive Fertigung—Anforderungen an qualitätsgesicherte Prozesse für additive Fertigungszentren

To Detect Defects Which Are Three-Dimensional Changes by Using Their bird's Eye View Images and Convolutional Neural Networks

Suzuki Ryuki and Haraguchi Harumi

Abstract The current understanding is that automating inspection is one of the most important issues in the manufacturing industry, and many studies are being conducted for automated inspection. Recently, there has been a lot of research on detecting defective products from images using neural networks. Previous research has focused on detecting only two-dimensional defects (flaws, chips, etc.) from images and not three-dimensional defects (warpage, overlap, etc.) using images. Additionally, the equipment required for 3D inspection tends to be expensive, large-scale, and costly, so the initial cost is high. Many manufacturing sites have not been able to introduce it. In this study, we perform the 3D inspection using industrial cameras introduced in many manufacturing sites. Specifically, we will detect 3D defects such as war pages of transport trays used in actual manufacturing sites from 2D images. First, since the captured image contains noise due to many shooting environments, appropriate image processing is applied to remove the noise. The two types of images used in this process are those taken at the actual manufacturing site and those taken in a laboratory with a good shooting environment. We took the tray images in the laboratory at different angles between 10 and 90 in 10° increments. The processed images are then inputted to transition trained CNNs (Convolutional Neural Networks) for deep learning to perform binary classification of abnormal and normal. We also use Grad-Cam to visualize the learning to understand which part of the image the network focused on for classification. As a result, it shows that the network was unsuccessful in performing binary classification on any birds-eye view images. In comparison, the Grad-Cam visualization results show that the network obtained candidate features for tray warping, a 3D change, from the 30 to 40° images for the images taken at the manufacturing site and the images taken in the laboratory.

Keywords Convolutional neural networks · Detection of three-dimensional defects · Automation system · Production management

S. Ryuki · H. Harumi (✉)
Ibaraki University, Hitachi City 316-8511, Japan
e-mail: harumi.haraguchi.ie@vc.ibaraki.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_16

1 Introduction

In our target manufacturing site, workers manually place transport trays one by one on the transport line, and they visually inspect the trays to remove defective products. The major defect in this section is the warped defect. Because of this defect is the easiest to fall off the transport line than other defects. However, it is also the defect that is most often missed by visual inspection. For this reason, automatic and accurate inspection is necessary, but due to the limitations of the transport line, it is not possible to install a 3D inspection machine. At best, it is possible to install a camera right above the transport line, so it is necessary to detect warpage, which is a three-dimensional change, using RGB overhead images.

Since the early 1970s, automatic visual inspection has been actively researched using images in various manufacturing sites [1]. Additionally, 3D inspection has been studied since that time, and the light section method and pattern projection method have been tried [2]. However, these methods tend to be large and expensive because they use machines dedicated to 3D inspection. Recently, there has been an active research on methods to perform end-to-end visual inspection using machine learning [3–5]. The automated inspection using machine learning can be divided into methods based on supervised learning and unsupervised learning. Some studies have also used machine learning to estimate 3D information in RGB images, but the accuracy is not high.

1.1 Supervised Learning

The supervised learning-based visual appearance inspection includes work on cosmetic bottles by Shiraishi [6] and cast parts by Tanaka et al. [7]. Shiraishi detected anomalies in scratches made on the surface of a metal bottle. They targeted anomalies that are simple to detect as they are well defined in the image. Tanaka et al. conducted an abnormality detection of casting products using images of challenges for visual inspection algorithms. They used simple supervised learning and semi-supervised learning to try to detect anomalies. The detection rate is 79% in supervised learning, and the false positive rate is 0.12%. In semi-supervised learning, where defect information is added, the detection rate is 96.55%, and the false-positive rate is 6.11%.

1.2 Unsupervised Learning

In unsupervised learning for visual inspection, Tachibana et al. have used images of industrial products [8], and Oshima et al. have used aircraft parts [9]. They used a method with a self-encoder that learns only typical images [10]. When a defective image was input into the encoder, some parts of the image cannot be recovered properly. So, anomaly detection is used in this case. The method is effective when no significant number of defective images are available, but it is difficult to find various defects. In this method, since the characteristics of the self-coder, it is necessary to set the parameters appropriately, and its accuracy depends on the operator.

1.3 3D Visual Inspection

3D information estimation from images has been studied. Vincent et al. have created struct2depth, which can estimate depth information and camera ego-motion from video images using unsupervised learning [11]. The idea is that a depth image is obtained without using the correct depth information. However, while there are examples of this technique for images taken outdoors, there are no examples for images used indoors, such as those used in inspections. The example of the outdoor image shows that only rough depth estimation is possible, which indicates that it is not suitable for inspections where accurate values are required.

2 Aim and Contribution of This Study

As described above, most of the research appearance inspection using images has been conducted on 2D defects or those for which 3D changes are clearly shown in the RGB image. However, it is not enough to study which 3D changes are neither clearly represented in the RGB image. This study prepares three kinds of images for appearance inspection of transport trays used in actual manufacturing sites. (1) bird's eye view images taken directly above the transport line at the actual manufacturing site using an industrial camera. (2) Images taken at different angles from 0° to 90° in a well-equipped laboratory. (3) Difference image between a quality tray image and a defective tray image using the images obtained in (1) and (2). Based on the (1), (2), and (3) images, binary classification is performed between normal and abnormal using CNNs with fine-tuned images. We also use Grad-Cam [12] to visualize the learning process to know which part of the image the network has focused on for classification. Therefore, we identify which part of the RGB image shows the warping of the tray. This study is organized as follows: 1. Description of the transport tray used in study 2. About image processing and Grad-cam 3. Binary classification experiments.

3 Transport Tray

We use three images for the experiment in this study. An RGB image taken on a transport line at the actual manufacturing site (Fig. 1a), an RGB image taken in a

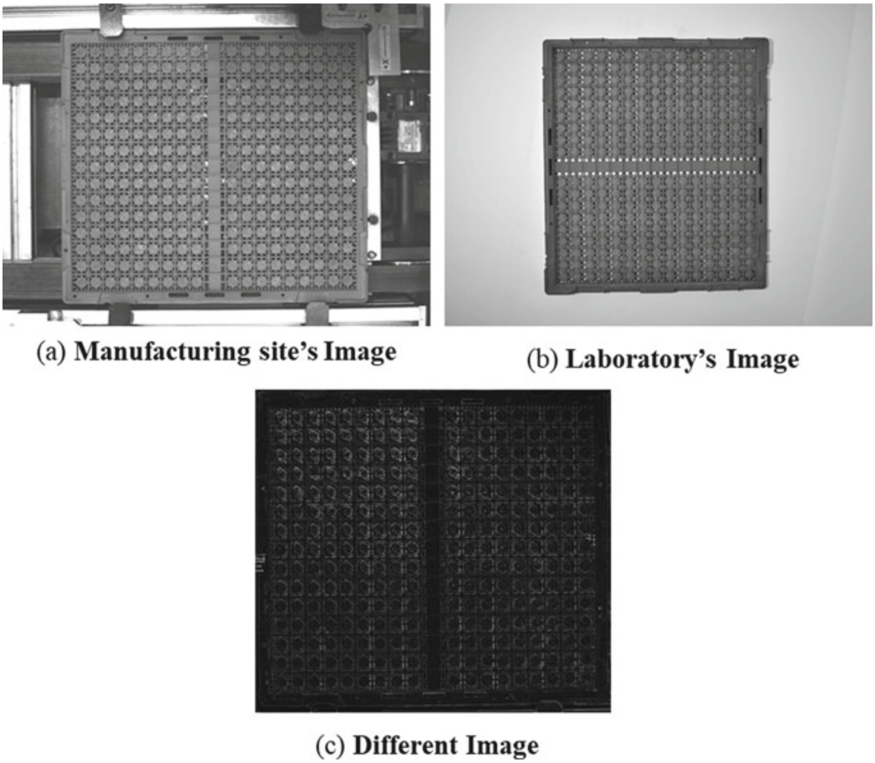


Fig. 1 Transport tray

laboratory where the shooting environment is well prepared (Fig. 1b), and a difference image created by combining these two images (Fig. 1c). The images were processed and cropped to emphasize the tray due to the influence of the hauling line and the shooting booth (Fig. 2).

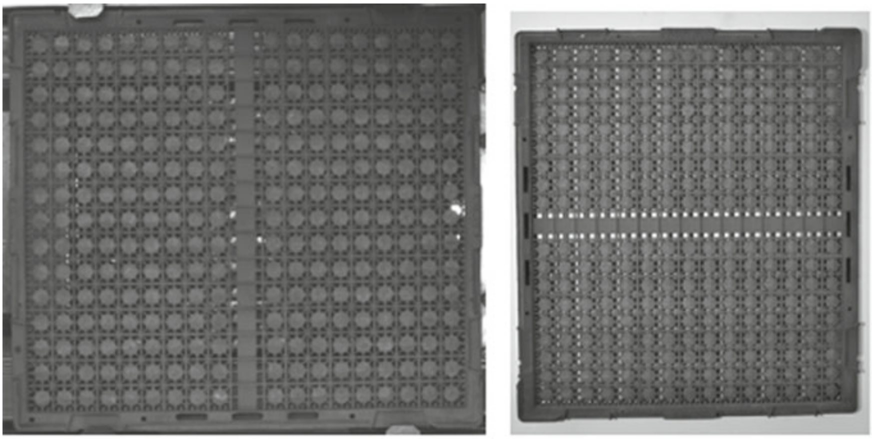


Fig. 2 After using image processing

4 Experiment

4.1 Method and Condition

In this experiment, we use three types of images for binary classification by the neural network. The three types of images are as follows; (1) images taken at a manufacturing site, (2) images taken in a laboratory, and (3) difference images. Each image is shown in Table 1. Each image is then visualized using Grad-CAM.

We use two neural networks. 1st is VGG16, and 2nd is ResNet50. These networks have already been trained in ImageNet [13]. The final layer of this network is modified to adapt to tray images to perform binary classification of images. The parameters of the neural networks used in the experiments are shown in Table 2. The evaluation of the binary classification results is based on three criteria: Accuracy, Precision, and Recall. The heat map image obtained by Grad-CAM is evaluated visually.

Table 1 Number of images

	(1) Image	(2) Image	(3) Image
True image (train)	54	12	50
False image (train)	6	6	10
True image (test)	35	12	30
False image (test)	6	6	11

Table 2 Parameter setting

Name	Parameters
Batch size	3
epoch	500
Optimizer	Stochastic gradient descent (SGD) (learning rate:1e-3, momentum:0.9)
Loss function	Cross entropy
Input image size	512*512(px)

4.2 Results and Discussion

Table 3 shows the respective evaluation values for the two types of images taken at the manufacturing site and the different images. Table 4 shows the evaluation values for the images taken in the laboratory at each angle. For Precision and Recall, we find an extreme difference between the defective image class and the normal image class regardless of the model. In the normal image class, the value is close to 1.0, but in the defective image class, the value is extremely low, around 0.1. The 50° image captured in the laboratory and classified by VGG16 has the best value, with Accuracy: 0.73, precision (defective class): 0.27, Recall (defective class): 0.39, which is not high. These results show that machine learning is mistakenly guessing many images as the normal product image class. In other words, it is not classifying the binary values correctly. Since the values are extremely different, the machine learning model may be causing the gradient loss of defective images (Fig. 3).

The heat map images created by each Grad-CAM are shown in Figs. 4 and 5. Figure 4 shows the evident changes output from (1) and (2). Figure 5 shows the output image of using VGG16. From a-1 to i-1 is the image of a quality tray. These images are using same tray, and it was taken the image of every 10 degrees changes an angle. From j-1 to r-1 is the image of a defective tray. Figure 6 shows the output image using ResNet50. From a-2 to i-2 is the image of a quality tray, and from j-2 to r-2 is the image of a defective tray.

We discuss the image taken at the manufacturing site: the VGG16 model reacts to the lower part of the outer frame of the tray. It is not a distorted part of the tray but a white stain on the tray. Similarly, ResNet50 reacted not to the distorted part

Table 3 The evaluation values for each of the two types of images

	Manufacturing site's image	Different image
Accuracy (True)	(0.96, 0.96)	
Accuracy (False)	(1.00, 0.96)	
Precision (True)	(0.09, 0.18)	
Precision (False)	(0.96, 1.00)	
Recall (True)	(0.09, 0.18)	
Recall (False)	(0.96, 0.96)	

Table 4 The evaluation values for the images taken in the laboratory at each angle

	10°	20°	30°	40°	50°	60°	70°	80°	90°
Accuracy (true)	(0.75, 0.75)	(0.75, 0.75)	(0.75, 0.75)	(0.75, 0.50)	(0.73, 1.00)	(0.75, 1.00)	(1.00, 1.00)	(1.00, 0.50)	(1.00, 0.50)
Accuracy (false)	(1.00, 0.75)	(0.75, 1.00)	(1.00, 1.00)	(1.00, 0.00)	(0.69, 1.00)	(0.75, 1.00)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
Precision (true)	(0.50, 0.00)	(0.00, 0.50)	(0.00, 0.50)	(0.50, 0.50)	(0.27, 1.00)	(0.00, 1.00)	(1.00, 0.00)	(0.00, 0.00)	(1.00, 0.00)
Precision (false)	(0.66, 1.00)	(1.00, 0.66)	(0.75, 0.66)	(0.66, 0.00)	(0.84, 1.00)	(0.00, 1.00)	(1.00, 0.00)	(1.00, 0.50)	(1.00, 0.50)
Recall (true)	(1.00, 0.00)	(0.00, 1.00)	(0.00, 1.00)	(1.00, 1.00)	(0.39, 1.00)	(1.00, 1.00)	(1.00, 1.00)	(0.00, 0.00)	(1.00, 0.00)
Recall (false)	(0.75, 0.75)	(0.75, 0.75)	(0.75, 0.75)	(0.75, 0.50)	(0.73, 1.00)	(0.75, 1.00)	(1.00, 1.00)	(1.00, 0.50)	(1.00, 0.50)

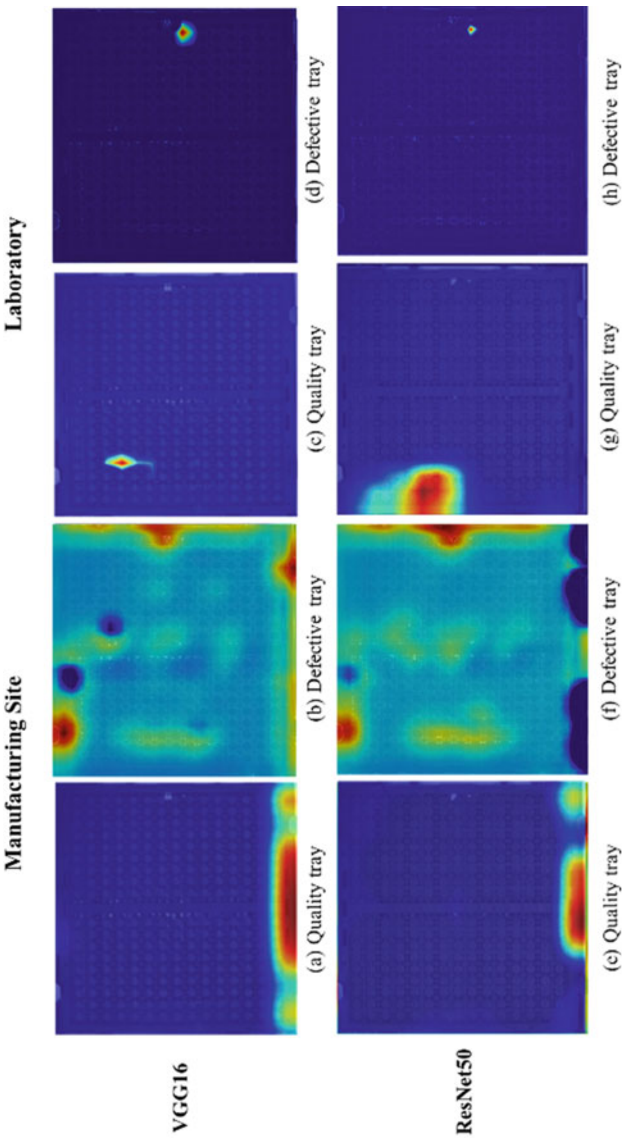


Fig. 3 Grad-CAM's heat map image (VGG16&ResNet50)

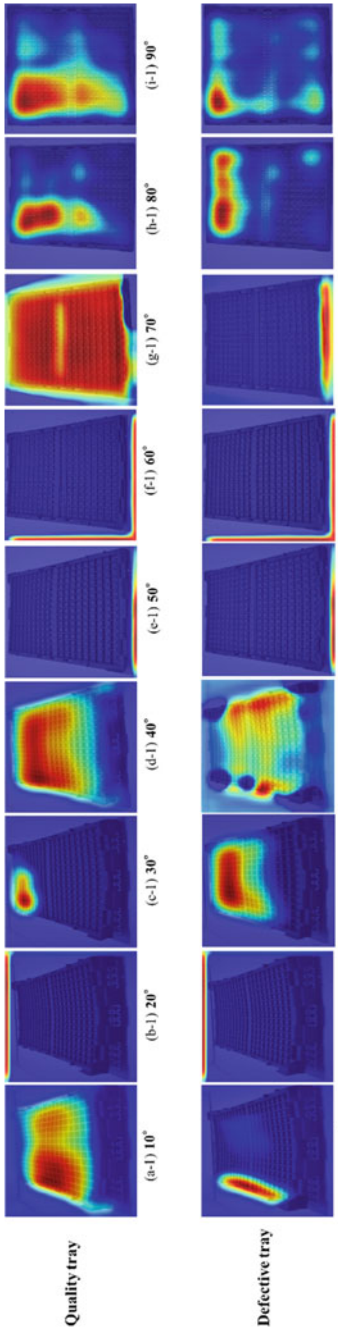


Fig. 4 Grad-CAM's heat map image (VGG16)

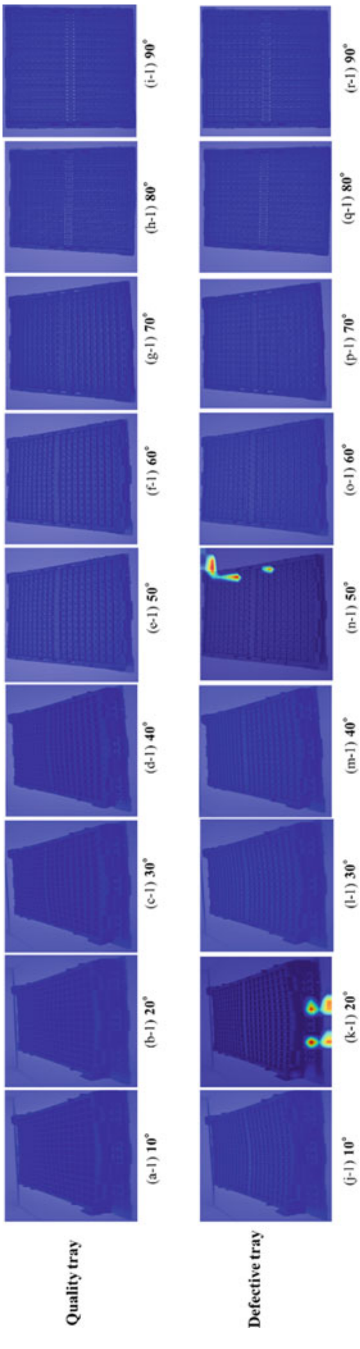


Fig. 5 Grad-CAM's heat map image (ResNet50)

of the tray but the dirty part of the tray. From the manufacturing RGB images, we detected no tray distortions. From the laboratory RGB images, ResNet50 did not produce results that captured the warping of the tray. In the VGG16 results, the same results were not obtained at 20, 50, and 60°. In the other angles of the image, we see areas of strong response. This area is most likely a distortion as it is not a stain on the tray. In the different images, ResNet50 did not produce a heatmap image with a strongly reacting area. The VGG16 image emphasizes a different location than the other two, and this location is also not a stain on the tray, so it is most likely pointing to a distorted part of the tray.

5 Conclusion

This study tried 3D inspection from 2D images using convolutional neural networks. Specifically, we want to detect the warpage trouble from transport trays used in actual manufacturing sites from 2D images. Since verifying the deference of angle, we took the sample images at various angles from 0° to 90° and executed computer experience. In the experiment, we got the warped candidate parts of the tray from the heat map images by Grad-CAM with VGG16 of two kinds of images. One is the image taken in the laboratory, and the other is the difference image among the three types of images. The numerical evaluation showed that machine learning was impossible to classify any image. The fact that most of the Grad-CAM heatmap images were blue on all sides and the values of Precision and recall were extremely indicated that gradient loss was occurring. Additionally, VGG16, which has a deeper layer than ResNet50, showed better results, suggesting that a model with a shallower layer is more appropriate for this image.

Our future studies aim to improve the number of image samples, the shooting environment and analyze the obtained heat map images. In the previous study, several hundred images were prepared. While in our study, the number of samples is small, about 50 at most. As in the following study, we should take several hundred images. Regarding the shooting environment, we had found that we got more meaningful heat map images when the images were taken in a laboratory with a good shooting environment than taking images in an actual manufacturing site. In this study, we found that the warpage of the tray was not clearly shown in the images, and features were minimal. Therefore, we should construct a noise-free shooting environment to make the warp of the tray stand out. Additionally, we should analyze in detail to understand the 3D changes in the RGB image and the match between the obtained heat map image and the actual warped parts. In the future, it is necessary to reflect the obtained analysis results in image processing to make the features more prominent.

References

1. Yada, T.: [II. Inspection by image processing] II. Gazousyori ni okeru kennsa (in Japanese). IEEJ J. Ind. Appl. **116**(1), 2–4 (1995). https://doi.org/10.1541/ieejias.116.2
2. Hata, S.: Trends of automatic visual inspection. J. Jpn. Soc. Precision Eng. **56**(8), 1366–1370 (1990). https://doi.org/10.2493/jjspe.56.1366
3. Hashimoto, D., Tamagaki, Y., Suitani, M., Nagata, T.: [<<No. 6>> machine learning and visual inspection: advancing automation of inspection and its challenges] <<No. 6>>Kikaigakusyu to gaikannkennsa: susumukennsa no zidouka to sonokadai (in Japanese). J. Soc. Instrum. Control Eng. **58**(9), 711–716 (2019). https://doi.org/10.11499/sicejl.58.711
4. Kubota, S.: Application of deep learning to visual inspection and research trends. J. Jpn Soc. Precision Eng. **85**(1), 27–30 (2019). https://doi.org/10.2493/jjspe.85.27
5. Nomura Y.: A review on anomaly detection techniques using deep learning. J. Soc. Mat. Sci. Jpn. **69**(9), 650–656 (2020). https://doi.org/10.2472/jsms.69.650
6. Shiraishi, Y.: An appearance inspection with defect size estimation based on deep learning algorithm. J. Surf. Finish. Soc. Jpn. **71**(7), 442–447 (2020). https://doi.org/10.4139/sfj.71.442
7. Tanaka, T., Kasahara, R.: Automated visual inspection technology using image data. J. Imaging Soc. Jpn. **55**(3), 348–354 (2016). https://doi.org/10.11370/isj.55.348
8. Tachibana, R., Matsubara, T., Uehara, K.: Anomaly manufacturing product detection using unregularized anomaly score on deep generative models. Kagoshima, p. 2A103 (2018). https://doi.org/10.11517/pjsai.JSAI2018.0_2A103
9. Ooshima, T., Suzuki, K., Maeda, M., Imai, S., Kimura, K.: Development of variational auto-encoder based anomaly detection technique for pattern images of honeycomb structure. J. IHI Technol. **60**(1), 45–51 (2020)
10. Kingma, D.P., Welling, W.: Auto-encoding variational bayes [cs, stat]. Available: http://arxiv.org/abs/1312.6114. Accessed 28 May 2021 [Online] (2014)
11. Casser, V., Pirk, S., Mahjourian, R., Angelova, A.: Depth prediction without the sensors: leveraging structure for unsupervised learning from monocular videos. AAAI **33**, 8001–8008 (2019). https://doi.org/10.1609/aaai.v33i01.33018001
12. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. Int. J. Comput. Vis. **128**(2), 336–359 (2020). https://doi.org/10.1007/s11263-019-01228-7
13. Deng, L., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

A Multi-Objective Fitness Function for Sequencing Robotic Assembly Operations with Deformable Objects Using a Genetic Algorithm with Constraint Satisfaction

Shir Ben-David and Sigal Berman

Abstract Constructing and optimizing an operation sequence is a major concern in production. However, designing high quality sequences is difficult and assembly sequence planning (ASP) is a NP-hard problem. Due to inherent process uncertainties production processes which include robotic manipulation of deformable objects, ASP can be even more complex than when only rigid objects are manipulated. Genetic algorithms (GA) are a commonly used heuristics for ASP. GA is suitable for ASP of robotic operations with deformable objects, which typically require addressing multiple objectives and which have complex production constraints. However, not all constraint may be explicitly known during the sequence design time. The current work examined ASP with different levels of known production process constraints. Integrating an arc consistency algorithm (AC3) for constraint satisfaction in the initial population generation process was implemented and compared to random population generation. Integrating process duration and the longest common subsequence (LCS) index (to sequences of similar products) in the fitness function was implemented and different integration methods were compared. Results show that the effects of using AC3 for initial population generation depend on chromosome length and are not related to the rate of addressed constraints. For long chromosomes AC3 based generation is considerably faster than random generation. The impact of adding LCS to the fitness function depends on the rate of constraints addressed and is not related to chromosome length. When not all the production constrains are addressed LCS increases the number of feasible solutions obtained and is not related to chromosome length.

Keywords Genetic algorithms · Arc consistency3 · Longest common subsequence

S. Ben-David · S. Berman (✉)
Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, 8410500 Beer-Sheva, Israel
e-mail: sigalbe@bgu.ac.il

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_17

1 Introduction

Assembly of deformable objects is complex since can require gentle handling and due to unintentional changes in object shape during manipulation which require motion adaptation. For human workers, such adaptations are typically trivial, however gentle handling and motion adaptation can be challenging for robotic systems [1, 2]. Constructing and optimizing assembly operation sequences is a major concern in production as the sequence affects both assembly time and production cost [3]. However, assembly sequence planning (ASP) is a NP-hard problem and therefore, designing high quality sequences is difficult [4]. The inherent uncertainty in robotic assembly of products with deformable objects makes their ASP even harder. ASP of products with deformable object has not been widely addressed in literature. A possible approach that can handle process uncertainty and complexity is to use a two-stage method, when suspected high quality sequences are generated using heuristics followed by testing with a physical simulation. Since simulations with physical engines have non-negligible run-time the thoroughness and quality of the first, heuristic stage is important [5].

Many heuristics have been developed for optimizing ASP, among them the biologically inspired genetic algorithms (GA) is a commonly used method. GA is loosely based on evolution theory and provides an iterative method for searching the solution space [4, 6]. Starting with an initial population of solutions, in each iteration the algorithm tries to improve the solutions using crossover and mutation functions and retains only the best solutions for the following iteration. GA offers a straightforward way to represent assembly sequences which can be iteratively optimized based on a fitness function. Among the advantages of GA are the straight-forward incorporation of fitness functions with multiple objectives and their flexibility in interpreting constraints [4]. These, make them suitable for ASP of robotic operations with deformable objects, which typically require addressing multiple objectives and which have complex constraints.

GA are greedy algorithms in terms of execution time and many iterations may be required for optimizing the solution. The initial population largely effects both convergence speed and the quality of the final solution [7]. Methods for generating a high-quality initial population that will converge faster towards the global optimum is an import research question [6, 8–10]. Assembly operations form interactions between objects, and are defined by the connection between the objects (glue, snap, etc.). Priority constraints detail which interaction must be handled first and proximity constraints requires contiguity between interactions. Zouita et al. [8] suggest integrating a constraint satisfaction problem (CSP) for the initial population generation. CSP is a powerful tool for modeling and solving combinatorial problems [8]. However, finding a solution to CSP is also NP-hard. Several filtering techniques have been developed for reducing the required search space [11]. Among the widely used techniques is Arc consistency [8, 11–13]. Arc consistency is suitable for handling binary constraints and guarantees that all the constraints regarding values of each variable in the solution domain can be satisfied by at least one value of the other variables [8]. The arc consistency3 (AC3) algorithm balances simplicity with efficiency as it maintains only a relatively small number of data structures during search [8, 11]. Another way to reduce execution time and improve quality is to look at similarity to previously found solutions. Such a resemblance can be measured by the longest common subsequence (LCS) index. LCS examines the relative order of actions yet, it does not mandate their consecutive positioning. This provides LCS the ability to inspect and gain insights for product variety.

The current work suggests a GA variant that applies constraint satisfaction with AC3 for initial population generation and integrates time duration, LCS, and constraint satisfaction in a multi-objective fitness function. The method is integrated with a database which stores information regarding the assembly process: products, assembly operations, assembly constraints, equipment, objects, and sequences of similar products. The database facilitates generation of feasible solutions based on both product assembly requirements and workcell capabilities. It additionally facilitates measuring similarity to sequences of similar products. Since not all constraint may be explicitly known during the sequence design time, different levels of known production process constraints are examined. We hypothesize that integration of AC3 will lead to faster generation of the initial chromosome population. We further hypnotized that integration of LCS with the fitness function will improve the quality of generated solution, depending on the percentage of specifically defined constraints.

2 Method

2.1 Environment

Nine products, which were variants of three product types (a smart light, a medical infusion kit, and a plastic lid) (Table 1) were tested. The algorithms were implemented using Python 3.7 with a PyCharm (JetBrains, The Netherlands) environment with a Windows 10 64-bit operating system running on an Intel Core i9-9900 K 3.60 GHz processor with 32 GB RAM.

2.2 Sequence Computation

For generating the initial population, relevant assembly operations are extracted from the database according to object material, available equipment, and the interaction type. The operations are divided into groups according to the interaction type and the number of interaction types defines the length of the chromosomes. One operation is randomly selected from each group and the set continues to the constraint satisfaction problem (CSP) algorithm to find a feasible location for each assembly operations in

Table 1 The tested product

Product	Variant	# Robotic arms	Components
Smart light	1	1	2 wires, 1 panel with 3 reflectors
	2	1	6 wires, 1 panel with 3 reflectors
	3	1	2 wires, 1 panel with 3 reflectors, 1 dimmer
Medical infusion kit	1	2	1 hose, 2 connectors
	2	2	1 hose, 2 connectors, 1 nozzle
	3	1	1 hose, 2 connectors
Plastic lid	1	2	1 square lid, 1 square gasket
	2	2	1 elliptical lid, 1 elliptical gasket
	3	1	1 square lid, 1 square gasket

the assembly sequence according to the constraints. Each chromosome of the initial population is a constraint satisfaction problem solution, and each gene represents a variable (assembly operation). The input to the constraint satisfaction algorithm is the chosen set of assembly operations and their priority and proximity constraints which detail the assembly ordering constraints. Here is an example of the formalization of the problem for three assembly operation (ao_1—ao_3): variables: {ao_1, ao_2, ao_3}, domains: {$d_i = \{1,2,3\}$ for $i = 1,2,3$}, and constraints: {$ao_1 < ao_3$ (priority constraint), $ao_1 + 1 = ao_2$ (proximity constraint)} meaning, ao_1 has to be assemble before ao_3 and ao_2 has be to assemble right after ao_1.

The constraint satisfaction algorithm begins with AC3 that is performed on the operations. Its objective is to lower the dimension of the possibilities for arranging the operations in the sequences and thus to simplify the problem. At the end of the algorithm, the problem of arranging the operations is solved maintaining arc consistency—for each value in the operation domain (possible location in sequence), there is some value in all other domains that satisfied this value (adheres to the productions constraints). The second algorithm is back tracking (BT). Its purpose is to deal with most situations where there is more than one value in operation's domains. Therefore, it is necessary to select a single value, i.e., a specific location for each operation. The choice of value for each operation should consider the production constraints but also provide a different value for each operation which constitutes its serial number within the assembly sequence. For each set of assembly operations that arrive to the CSP algorithm, each BT solution produced creates one chromosome. Initial population generation will continue until there are in chromosomes in the population described in Fig. 1.

Once the initial population is found according to the assembly constraints, the GA is iteratively executed with a fitness function that integrates production time (calculated in Eq. 1), which is a classical assembly performance measure for assembly sequence planning, and resemblance to existing assembly sequences of products in the same field. The resemblance index implemented using LCS, compares the new generate sequence, to similar sequences of products from the same field that have been

GA initial population generation with CSP

Input: Assembly operations for product, assembly constraints

Output: n feasible chromosomes

```
1. AOs = all  possible assembly operations of the product
2. L = number of interactions in the product.
3. While number of chromosomes in the population < n
4.    AO_set = select  L assembly operations from AOs (one for each
      interaction).
5.    set CSP: arcs, domains, constraints ,According to AO_set.
6.    new_domains = AC3(input= arcs, domains, constraints)
7.    new chromosome = BT(input= arcs, new_domains, constraints)
8.       If {find feasible new chromosome}
9.          If {new chromosome is not identical to member in the
               population}
10.            Add new chromosome to population
11.         End if
12.      End if
13. End while
```

Fig. 1 Pseudo code for initial population generation with CSP

validated and tested. The resemblance is calculated according to the connection type of each assembly operation in the sequence (connection type between the components in the operation), i.e., the connection type of each assembly operation and its location in the sequence are compared to those of the validated sequences. For each possible chromosome c_i, calculated the LCS index to all similar chromosomes c_j in the same field. The input to the lcs is two lists with the connection type of the assembly operations in the sequence in the relative order. For lcs (c_i, c_j) c_i has length m, and c_j has length n, where i, j denote the place in the chromosome $0 \leq i \leq m$, $0 \leq j \leq n$ and A is the array for storing the iterations values as in Eq. 2. The value of the resemblance index is determined to be the maximum similarity of the tested sequence against all validated sequences in the field (Eq. 3). The fitness function combines the two metrics, production time and LCS using a weight parameter α as shown in Eq. 4 in order to maximize the value and the quality of the solutions.

$$PD(c_i) = \frac{\frac{1}{\sum_{j=1}^{k} T_j}}{\max \frac{1}{\left(\sum_{j=1}^{k} T_j\right)}} \quad \text{k, the length of } c_i,\ T_j \text{ the duration of operation j in } c_i \tag{1}$$

$$lcs\big(c_i[m], c_j[n]\big) = \begin{cases} A[i, j] = 0 & if\ i = 0\ or\ j = 0 \\ A[i, j] = A[i-1, j-1] + 1 & if\ i, j > 0\ and\ c_i[i] = c_j[i] \\ A[i, j] = \max((A[i-1]), (A[j-1])) & otherwise \end{cases} \tag{2}$$

Table 2 Fitness function methods

	h1	h2	h3	s1	s2	s3
Percentage of constraints (%)	100	100	100	60	60	60
Fitness function	LCS: threshold Duration = 1	LCS = 0.5 Duration = 0.5	Duration	LCS: threshold Duration = 1	LCS = 0.5 Duration = 0.5	Duration

$$LCS(c_i) = \text{Max}\big(lcs\big(\text{c}_\text{i}, c_j\big)\big)\ c_j \in \{\text{set of similar chromosomes in the same field}\} \tag{3}$$

$$FC(C_i) = \alpha * PD(C_i) + (1 - \alpha) * LCS \tag{4}$$

2.3 Experimental Procedure

For testing the generation of the initial population with CSP we compared the results to random generation [8]. Twenty repetitions were performed for each method and variant, and three rates of assembly constraints were used, 100, 60, 20% (20·2·9·3 = 1080 runs). The size (n) of the required initial population was set to 10 in all runs and the time was limited to 600 s (10 min). The weight parameter was set to 0.5 in all the runs.

The initial population for testing the fitness function was based on the initial population created using AC3 with the suitable percentage of constraints. For each constraint level, three fitness functions were defined according to the relation between production time and LCS in the fitness function. The relations tested (Table 2) included (1) threshold based on chromosomes with LCS index above 50%, and quality computed based on process duration (2) Quality based on equal wieghts of process duration and LCS (3) Quality based on process duration, no regard to LCS.

2.4 Analysis and Statistical Analysis

For testing the generation of the initial population, we checked the number of variants for which the initial population converged to a solution faster with the CSP than with random generation. We only tested an observation that converged to a solution. We computed the two-sample-t confidence interval (Welch) for independent samples with a two-sided hypothesis. The null hypothesis was that the initial population run-time is equal in both generation methods: csp and random. We tested the hypotheses with a confidence interval of 0.95.

For testing the multi objective fitness function, we measured the number of feasible sequences out of the total sequences produced and converged for each variant, and we measured the processing duration for each variant and fitness function. We performed an ANOVA test for each variant to examine the difference in process duration between the different fitness functions. We also tested the Bernoulli distribution for feasible sequences when not all the assembly constraints were given.

3 Results

3.1 Initial Population Generation

The results of the initial population generation are summarized in Table 3. For the smart light product (chromosome length relatively long) we rejected the null hypothesis for any percentage of constraints and found that the initial population with the CSP converges faster. For variant 2 and 3 of the smart light product, the random initial population generation could not find ten feasible chromosomes in 600 s and therefore did not converge to even one solution. For variant 2 of the smart light (the longest one) none of the observations were able to produce a single feasible chromosome in the given time for any constraint percentage. However, for variant 3, about 10 observations produced 1–3 feasible chromosomes (but not 10) in the given time. For the medical infusion kit (medium chromosome length), for some constraint percentages we reject the null hypothesis (csp run-time was shorter), and for some the differences were not significant. For the plastic lid product (short chromosome length) we do not reject the null hypothesis, except for constraints percentage of 60,20 in variant 3, in which once SP run-time is shorter and in once random is shorter. To summarize, we examined the effect of each of the factors: chromosome length, algorithm (random/CSP), and percentage of constraints versus time of initial population generation. The length of the chromosome, the algorithm, and the interaction between the length of the chromosome and the algorithm were found to be significant (significant code ***) but the percentage of constraints was found to be insignificant.

3.2 Multi Objective Function

The results of examining the multi objective function are summarized in Table 4. For the probability of feasible sequences, when 100% of the assembly constraints were given to the problem, as expected, all sequences were feasible. The results for the case when not all assembly constraints were given the problem varied, for s1 the percentage of the feasible sequences out of the total converged sequences was 0.501, for s2 0.705 and for s3 0.511. It can be seen that in terms of generating feasible sequences method s2 is preferable. The threshold value of the LCS may have been

Table 3 Initial population experiment results

Product	Variant (Chromo-some length)	Num of converged observations CSP	Num of converged observations random	Mean initial population run-time CSP	Mean initial population run-time random	P-value
Smart light	1 (12)	20, 20,20	19, 18,19	**10.025**, **9.947**, **11.107**	430.348, 405.936, 431.812	***, ***, ***
	2 (32)	20,15,20	0,0,0	27.950, 207.697, 79.467	–	–
	3 (13)	20,20,20	0,0,0	11.143, 10.811, 10.763	–	–
Medical infusion kit	1 (7)	20,20,20	20,20,20	5.950, 5.882, **5.784**	6.060, 5.893, 5.891	ns, ns, **
	2 (8)	20,20,20	20,20,20	**6.701**, **6.719**, 6.730	6.905, 6.843, 6.816	***, ***, ns
	3 (7)	20,20,20	20,20,20	5.973, 5.838, 5.822	6.093, 5.847, 5.925	0.06, ns, ns
Plastic lid	1 (5)	20,20,20	20,20,20	4.210, 4.184, 4.292	4.183, 4.167, 4.118	ns, ns, ns
	2 (5)	20,20,20	20,20,20	4.198, 4.185, 4.182	4.239, 4.193, 4.197	ns, ns, ns
	3 (5)	20,20,20	20,20,20	4.144, 4.166, **4.209**	4.164, **4.095**, 5.018	ns, ***, ***

Each cell summarized the results for the three percentages of constraints in the following order 100%, 60%, 20%. In bold the significant values, *** < 0.001, ** < 0.01, ns = not significant.

too low and most of the sequences were above it. Therefore, the use of the LCS threshold not meaningful. However, when the LCS index was incorporated into the fitness function as in s2, a higher number of feasible sequences was achieved. The results from the ANOVA of the process duration indicated significance for the medical infusion kit. The process duration of h1–h3 methods (100% constraints) was shorter.

Table 4 Fitness function experiment results

Product	Variant	h1	h2	h3	s1	s2	s3	Anova
Smart light	1	20/20,151.7, 3.8	20/20,151.8, 3.5	20/20,152.5, 3.3	3/20, 154, 2.6	3/20,151.6, 1.5	5/20,156.6, 4.1	ns
	2	20/20, 449.1,6.1	20/20, 448.6, 5.8	20/20,445.9, 6.4	1/16,462, -	0/17, -, -	0/16, -, -	-
	3	20/20,173.7, 3.2	20/20,173, 3.8	20/20,173.8,4.2	14/20,173.5, 3	11/20,174.9, 4.6	15/20,175.4, 4.1	ns
Medical infusion kit	1	20/20,**42.1**, 1.8	20/20,**41.8**, 1.7	20/20,**42**, 2	9/20,**43.6**, 2.7	18/20,**44.8**,2.6	10/20,**44.1**,2.6	***
	2	20/20,**54**, 2.7	20/20,**53.5**, 3.4	20/20,**52.3**,1.8	7/20,**54.8**, 3.8	17/20,**58.2**, 5.4	9/20,**54.1**, 2.4	***
	3	20/20, **56.5**,1.5	20/20,**56.3**, 1.1	20/20,**56.8**,1.5	12/20, **57.6**,1.6	18/20,**58.5**, 2.8	10/20,**57.2**, 1.3	**
Plastic lid	1	20/20,51,2	20/20,52.3, 3.2	20/20,51.5,2.7	15/20,51.8, 3.1	20/20,52.8, 3.6	14/20,52.1, 2.3	ns
	2	20/20,51, 2.6	20/20,51.6, 3.1	20/20,51.4,3.4	14/20,51.9,3.7	20/20,52.9, 4.1	13/20,53.6, 3.2	ns
	3	20/20,63.1, 0.4	20/20,63.1, 0.4	20/20,63.3,0.6	15/20,63.2,0.7	20/20,63.7, 1.2	16/20,63.1, 0.5	0.09

Each cell includes the number of final feasible observations/the number of converged observations, the mean process duration, and the duration standard deviation. *** < 0.001, ** < 0.01, ns = not significant.

4 Conclusions

The effects of using AC3 for initial population generation depend on chromosome length and are not related to the rate of presented constraints. For long chromosomes AC3 based generation is considerably faster than random generation. For short chromosomes using AC3 is equivalent to random generation. The impact of adding LCS to the fitness function depends on the rate of constraints presented and is not related to chromosome length. When not all the production constrains are presented, LCS increases the number of feasible solutions obtained and is not related to chromosome length. When all the production constraints are presented, addition of LCS does not affect the number of feasible solutions obtained. These findings can be used to direct sequence generation based on product characteristics and the availability of the production process constraints.

Acknowledgements This research was supported by the Israel Innovation Authority as part of the Assembly by Robotic Technology (ART) consortium [grant number 67436].

References

1. Ghandi, S., Masehian, E.: Assembly sequence planning of rigid and flexible parts. J. Manuf. Syst. **36**, 128–146 (2015). https://doi.org/10.1016/j.jmsy.2015.05.002
2. Sanchez, J., Corrales, J., Bouzgarrou, B., Mezouar, Y.: Robotic manipulation and sensing of deformable objects in domestic and industrial applications: a survey. Int. J. Robot. Res. **37**, 688–716 (2018). https://doi.org/10.1177/0278364918779698
3. Mishra, A., Deb, S.: Robotic assembly sequence planning and optimization by cuckoo search algorithm. DEStech Trans. Eng. Technol. Res. (2018). https://doi.org/10.12783/dtetr/icpr2017/17678
4. Abdullah, M., Ab Rashid, M., Ghazalli, Z.: Optimization of assembly sequence planning using soft computing approaches: a review. Arch. Comput. Methods. Eng. **26**, 461–474 (2018). https://doi.org/10.1007/s11831-018-9250-y
5. Ben-David, S., Shneor, R., Zuler, S., Mann, Z., Greenberg, A., Berman, S.: Simulation-based two stage sequencing of robotic assembly operations with deformable objects. In: INCOM Conference (2021)
6. Marian, R., Luong, L., Abhary, K.: A genetic algorithm for the optimisation of assembly sequences. Comput. Ind. Eng. **50**, 503–527 (2006). https://doi.org/10.1016/j.cie.2005.07.007
7. Lin, L., Gen, M.: Hybrid evolutionary optimisation with learning for production scheduling: state-of-the-art survey on algorithms and applications. Int. J. Prod. Res. **56**, 193–223 (2018). https://doi.org/10.1080/00207543.2018.1437288
8. Zouita, M., Bouamama, S., Barkaoui, K.: Improving genetic algorithm using arc consistency technic. Proc. Comput. Sci. **159**, 1387–1396 (2019). https://doi.org/10.1016/j.procs.2019.09.309
9. Deng, Y., Liu, Y., Zhou, D.: An improved genetic algorithm with initial population strategy for symmetric TSP. Math. Probl. Eng. **2015**, 1–6 (2015). https://doi.org/10.1155/2015/212794
10. Vlašić, I., Ðurasević, M., Jakobović, D.: Improving genetic algorithm performance by population initialisation with dispatching rules. Comput. Ind. Eng. **137**, 106030 (2019). https://doi.org/10.1016/j.cie.2019.106030

11. Li, H.: Narrowing support searching range in maintaining arc consistency for solving constraint satisfaction problems. IEEE Access. **5**, 5798–5803 (2017). https://doi.org/10.1109/ACCESS.2017.2690672
12. Wang, R., Yap, R.: Arc consistency revisited. Springer, Cham. 11494 (2019). https://doi.org/10.1007/978-3-030-19212-9_40
13. Christian Bessiere, R.: Some practicable filtering techniques for the constraint satisfaction problem. http://citeseerx.ist.psu.edu/viewdoc/summary. 10.1.1.54.4126

A State Tracing Method for the System Data Obeying Poisson Distribution

Shuto Tanabashi, Yuto Torisaka, Yasuhiko Takemoto, and Ikuo Arizono

Abstract In the case that abnormality in a state of systems is found, identifying when the state in the system changed, that is, identifying the change-point is helpful to investigating the cause of the system abnormality and dealing with it. In our research, a method of identifying change-points and tracing the system state has been discussed based on the data from the system, when the characteristic values representing the system state obey the Poisson distribution. Specifically, based on the likelihood theory and information statistics, a method of selecting the best statistical model for explaining the state changes of the system has been proposed as the state transition tracing method. Then, the effectiveness of the proposed state tracing method has been verified through numerical simulation. In addition, an application example of inferring the cause of state changes from the outcome of our proposed method has been shown.

Keywords Change point detection · Information criterion · Poisson distribution · Maximum likelihood method

1 Introduction

When abnormality is detected in a system, identifying the point in time that the abnormality occurred is useful for investigating the cause of the system abnormality. The method of determining the point of occurrence of abnormality is generally called change-point detection, and various studies have been practiced on change-point detection of process condition in production systems [1–4]. For example, Samuel

S. Tanabashi (✉) · Y. Torisaka · I. Arizono
Okayama University, Okayama, Japan
e-mail: tanabashi@s.okayama-u.ac.jp

Y. Torisaka
e-mail: pkcc48um@s.okayama-u.ac.jp

Y. Takemoto
Kindai University, Osaka, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_18

and Pignatiello [2] have proposed a method of estimating the point that the Poisson rate parameter has changed in a system based on the maximum likelihood theory, in the situation that the system state is described by a Poisson distribution. These studies on the change-point detection based on the maximum likelihood theory have supposed the situation where the change of state occurs at just once.

However, it is possible that the state change has occurred multiple times instead of just once until abnormality is detected. Hence, a method of estimating the state change-points considering the possibility that the state changes multiple times before abnormality in the system is detected might be necessary. In such cases, although the maximum likelihood theory might detect change-points if the number of system changes is specified, else if, it is not suitable for simultaneously estimating the number of system changes and detecting change-points. Therefore, in the case that the number of system changes is unknown, the method of identifying the number of change-points is required in addition to the maximum likelihood theory.

Hence, Takemoto and Arizono [5] have investigated a method of estimating the state change-points considering the possibility that the state changes multiple times before abnormality in the system is detected. In Takemoto and Arizono [5], the Akaike Information Criterion (AIC) [6], expanding the concept of the maximum likelihood theory, has been applied in order to construct a state transition tracking method of detecting the multiple state change-points. In concrete, the state transition tracking method of detecting the multiple state change-points about mean and variance in the case that the system state is described by normal distribution has been proposed [5]. Note that AIC considers the number of parameters describing the statistical model in addition to the goodness of fit between the data and the statistical model regarding the likelihood.

This paper discusses a state transition tracking method of extracting changes in the state of a system based on the observed data of characteristic values in a situation where characteristic values obey a Poisson distribution. Note that this study deals with the situation where there may be multiple change-points in the rate parameters of the Poisson distribution. Thereby, based on the concepts of maximum likelihood theory and information statistics, we propose a state transition tracking method of selecting the most preferred statistical model for the temporal state change of a system characterized by a Poisson process. In addition, the effectiveness of the proposed method is verified through several numerical examples.

2 Change-Point Detection in a Single State Change

First, we explain the maximum likelihood method of estimating the Poisson rate parameter in the case of the change-point detection under the assumption of a single state change for the Poisson distribution (See Samuel and Pignatiello [2]). When the observed data x follows the Poisson distribution of parameter λ, the probability mass function of the Poisson distribution of parameter λ is given by

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{1}$$

Under this Poisson distribution assumption, suppose that each data $x_i, i = 1, \cdots, n$ is obtained by the i th observation. In this case, the log-likelihood function is given as

$$\log L = \sum_{i=1}^{n} \{x_i \log \lambda - \lambda - \log(x_i!)\} \tag{2}$$

From this first-order condition on the log-likelihood function in Eq. (2), the maximum likelihood estimator for λ is obtained as follows:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3}$$

Now, we assume a situation where the Poisson rate parameter of the initial state is known as $\lambda = \lambda_0$ and the state changes just once. Such a situation is assumed as the following statistical model:

$$\lambda = \begin{cases} \lambda_0, & i = 1, \cdots, \tau, \\ \lambda_1, & i = \tau + 1, \cdots, T, \end{cases} \tag{4}$$

where λ_1 means the Poisson rate parameter after the state change, τ is the point of state change, and both of them are unknown. In addition, T is the number of observations. By fitting Eq. (3) to the statistical model in Eq. (4), the maximum likelihood estimator of $\hat{\lambda}_1$ is expressed by

$$\hat{\lambda}_1 = \frac{\sum_{i=\tau+1}^{T} x_i}{T - \tau} \tag{5}$$

Therefore, the maximum log likelihood for λ_1 is given as

$$\begin{aligned} \log L &= \sum_{i=1}^{\tau} (x_i \log \lambda_0 - \lambda_0 - \log(x_i!)) + \sum_{i=\tau+1}^{T} (x_i \log \hat{\lambda}_1 - \hat{\lambda}_1 - \log(x_i!)) \\ &= \sum_{i=\tau+1}^{T} \left\{ x_i \log \frac{\hat{\lambda}_1}{\lambda_0} - \left(\hat{\lambda}_1 - \lambda_0\right) \right\} + C. \end{aligned} \tag{6}$$

Remark that C is constant as

$$C = \sum_{i=1}^{T} \{x_i \log \lambda_0 - \lambda_0 - \log(x_i!)\} \tag{7}$$

Furthermore, the maximum likelihood estimator of τ can be defined as follows:

$$\hat{\tau} = \arg \max_{1 \leq \tau \leq T} \sum_{i=\tau+1}^{T} \left\{ x_i \log \frac{\hat{\lambda}_1}{\lambda_0} - \left(\hat{\lambda}_1 - \lambda_0 \right) \right\} \tag{8}$$

In the statistical model described as Eq. (4), the estimators of the Poisson rate parameter and the change-point are obtained based on Eqns. (5) and (8).

3 A Method of Tracking State Transition

In this study, we consider the state transition tracking method in the situation of supposing multiple change-points for the Poisson rate parameter. We assume that the observed data x_i up to the first change-point follows a Poisson distribution with a specified parameter λ_0. Further, assume the Poisson rate parameter has changed K times until reaching the observation point T and λ_k indicates the Poisson rate parameter after the $(k-1)$ th change-point. In this case, the statistical model for the Poisson rate parameter change process in the observation point T is expressed as follows:

$$\boldsymbol{\lambda}^{(K)} = \begin{cases} \lambda_0, i = 1, \cdots, \tau_0, \\ \lambda_1, i = \tau_0 + 1, \cdots, \tau_1, \\ \vdots \\ \lambda_k, i = \tau_{k-1} + 1, \cdots, \tau_k, \\ \vdots \\ \lambda_K, i = \tau_{K-1} + 1, \cdots, \tau_K (= T). \end{cases} \tag{9}$$

Note that $\tau_0, \cdots, \tau_k, \cdots, \tau_{K-1}$ denote the change-points for the Poisson rate parameter, and $\lambda_1, \cdots, \lambda_k, \cdots \lambda_K$ describe the Poisson rate parameters after change of the system state at $\tau_0, \cdots, \tau_k, \cdots, \tau_{K-1}$. Note that λ_0 is the specified initial Poisson rate and $\tau_k = T$ is the terminal point in time. $\boldsymbol{\lambda}^{(K)} \equiv (\lambda_1, \cdots, \lambda_k, \cdots, \lambda_K)$, $\boldsymbol{\tau}^{(K)} \equiv (\tau_0, \cdots, \tau_k, \cdots, \tau_{K-1})$ and the total change number K are the model description parameters. And then, since these parameters $\boldsymbol{\lambda}^{(K)}$ and $\boldsymbol{\tau}^{(K)}$ are unknown, so they should be estimated in the identification for the statistical model. That is, our problem of the identification for the statistical model in Eq. (9) is defined as the optimum estimation problem for the unknown parameters K, $\boldsymbol{\tau}^{(K)}$, and $\boldsymbol{\lambda}^{(K)}$.

First, by fixing K and $\tau^{(K)}$, consider to estimate the Poisson rate parameters $\boldsymbol{\lambda}^{(K)}$. The log-likelihood function $\ell(\boldsymbol{\lambda}^{(K)})$ for the Poisson rate parameters $\boldsymbol{\lambda}^{(K)}$ under the statistical model in Eq. (9) prescribing K and $\boldsymbol{\tau}^{(K)}$ is given by

$$\ell(\boldsymbol{\lambda}^{(K)}) = \sum_{k=0}^{K} \sum_{i=\tau_{k-1}+1}^{\tau_k} \{x_i \log \lambda_k - \lambda_k - \log(x_i!)\} \tag{10}$$

where $\tau_{-1} \equiv 0$. Based on Eq. (10), under the fixed parameters K and $\tau^{(K)}$, the maximum likelihood estimator of λ_k can be obtained as

$$\hat{\lambda}_k = \frac{1}{\tau_k - \tau_{k-1}} \sum_{i=\tau_{k-1}+1}^{\tau_k} x_i, \;\; k = 1, \cdots, K. \tag{11}$$

In this case, under the set of maximum likelihood estimators of rate $\hat{\boldsymbol{\lambda}}^{(K)} = (\hat{\lambda}_1, \cdots, \hat{\lambda}_k, \cdots, \hat{\lambda}_K)$ given by Eq. (11), the set of maximum likelihood estimators $\hat{\boldsymbol{\tau}}^{(K)} = (\hat{\tau}_0, \cdots, \hat{\tau}_k, \cdots, \hat{\tau}_{K-1})$ of the change-points $\boldsymbol{\tau}^{(K)}$ are given to satisfy the following relation:

$$\hat{\boldsymbol{\tau}}^{(K)} = \arg\max_{\boldsymbol{\tau}^{(K)}} \ell(\hat{\boldsymbol{\lambda}}^{(K)}) \tag{12}$$

Note that we can transform the log-likelihood function as follows:

$$\begin{aligned} \ell(\hat{\boldsymbol{\lambda}}^{(K)}) &= \sum_{k=0}^{K} \sum_{i=\tau_{k-1}+1}^{\tau_k} \left\{x_i \log \hat{\lambda}_k - \hat{\lambda}_k - \log(x_i!)\right\} \\ &= \sum_{k=1}^{K} \sum_{i=\tau_{k-1}+1}^{\tau_k} \left\{x_i \log \frac{\hat{\lambda}_k}{\lambda_0} - \left(\hat{\lambda}_k - \lambda_0\right)\right\} + \sum_{i=1}^{T} \{x_i \log \lambda_0 - \lambda_0 - \log(x_i!)\} \\ &= \sum_{k=1}^{K} \sum_{i=\tau_{k-1}+1}^{\tau_k} \left\{x_i \log \frac{\hat{\lambda}_k}{\lambda_0} - \left(\hat{\lambda}_k - \lambda_0\right)\right\} + C. \end{aligned} \tag{13}$$

Therefore, the optimization problem in Eq. (12) can be represented as the following dynamic problem (DP) as follows:

$$\max_{\boldsymbol{\tau}^{(K)}} D_K(\boldsymbol{\tau}^{(K)}) = \max_{\tau_{K-1}} \left\{ \sum_{i=\tau_{K-1}+1}^{\tau_K} \left\{x_i \log \frac{\hat{\lambda}_K}{\lambda_0} - \left(\hat{\lambda}_K - \lambda_0\right)\right\} + \max_{\boldsymbol{\tau}^{(K-1)}} D_{K-1}(\boldsymbol{\tau}^{(K-1)}) \right\}, \tag{14}$$

where

$$\begin{cases} D_0(\boldsymbol{\tau}^{(0)}) \equiv 0, \\ D_m(\boldsymbol{\tau}^{(m)}) \equiv \sum_{k=1}^{m} \sum_{i=\tau_{k-1}+1}^{\tau_k} \left\{ x_i \log \frac{\hat{\lambda}_k}{\lambda_0} - \left(\hat{\lambda}_k - \lambda_0 \right) \right\} \\ \qquad = \sum_{i=\tau_{m-1}+1}^{\tau_m} \left\{ x_i \log \frac{\hat{\lambda}_m}{\lambda_0} - \left(\hat{\lambda}_m - \lambda_0 \right) \right\} + D_{m-1}(\boldsymbol{\tau}^{(m-1)}), \quad m = 1, \cdots, K. \end{cases} \tag{15}$$

Remark that $\hat{\boldsymbol{\tau}}^{(m)}$, $m = 0, \cdots, K-1$, means $\hat{\boldsymbol{\tau}}^{(m)} = (\hat{\tau}_0, \cdots, \hat{\tau}_m)$, and the parameters $\tau_{m+1}, \cdots, \tau_{K-1}$ are given outside the stage m for determining τ_m in DP formulated like Eq. (14).

As the above, the problem of estimating the parameters K, $\boldsymbol{\tau}^{(K)}$ and $\boldsymbol{\lambda}^{(K)}$ can be specified based on the maximum likelihood theory. However, if the fitness of the statistical model to the data is simply considered, the statistical models with the large number of change-points are preferred, that is, the number of change-points might be equal to the number of observations. This result clearly leads to overfitting problems. Therefore, with the intention of solving this problem, we consider that the Akaike Information Criterion (AIC) should be employed as the criterion for the preferred model selection. By using AIC, we can select adequately the statistical model by considering the balance between the likelihood for the data and the number of unknown parameters. In concrete, AIC for the statistical model selection under consideration is defined as follows:

$$AIC(K) = -2\ell(\hat{\boldsymbol{\lambda}}^{(K)}) + 2B_K, \tag{16}$$

where B_K indicates the number of model parameters to be estimated in the statistical model. In this case, since the statistical model has been defined by $\boldsymbol{\tau}^{(K)}$ and $\boldsymbol{\lambda}^{(K)}$, B_K is given as

$$B_K = 2K \tag{17}$$

As a consequence, based on AIC, the number K of the change-points can be decided as

$$\hat{K} = \arg\max_K AIC(K) \tag{18}$$

4 Numerical Verification

To verify the effectiveness of the proposed state transition tracking method, we apply it to the analysis of time series data on the Poisson random numbers generated under a scenario defined by some Poisson rates and change-points, and then confirm that

Table 1 Scenario for verification

λ		τ	
λ_0	20	τ_0	8
λ_1	5	τ_1	19
λ_2	25	τ_2	24
λ_3	35	τ_3	40
λ_4	15	τ_4	50

Table 2 Estimated values by numerical verification

$\hat{\lambda}$		$\hat{\tau}$	
λ_0	20.0	$\hat{\tau}_0$	8
$\hat{\lambda}_1$	5.55	$\hat{\tau}_1$	19
$\hat{\lambda}_2$	25.2	$\hat{\tau}_2$	24
$\hat{\lambda}_3$	34.0	$\hat{\tau}_3$	40
$\hat{\lambda}_4$	14.3	τ_4	50

the estimated Poisson rate follows the state change as in the scenario. Table 1 shows the scenario used in this numerical verification. We set the number of observations to $T = 50$ and the number of changes to $K = 4$. The analysis results for Poisson random numbers generated under the scenario defined in Table 1 are shown in Table 2 and Fig. 1.

From Fig. 1, the estimated the Poisson rate parameters seem to fit the data well. Through the comparison between Tables 1 and 2, it can be seen that all the change-points are estimated accurately. The increase and decrease of the Poisson rate parameter at each change-point are also estimated accurately. These results indicate that

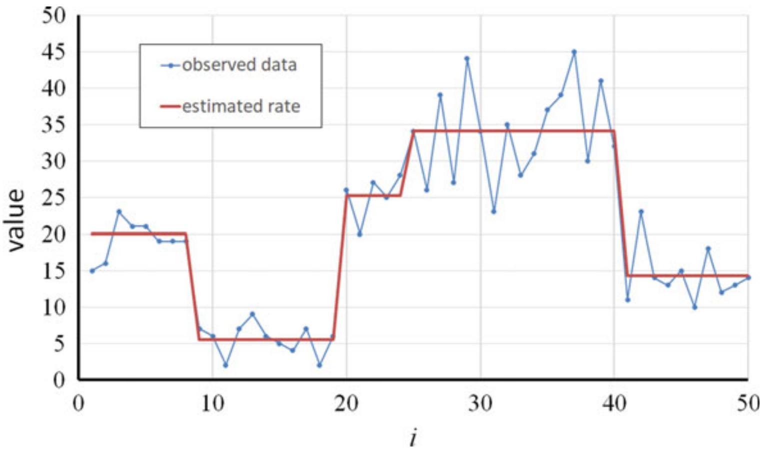


Fig. 1 Numerical simulation

the proposed method of state tracking is effective in extracting the change-points of state changes and inferring the point in time when the cause of the change should be investigated.

5 Conclusion

It is an important issue in system management that the causes of state changes are inferred and verified based on the state change-points obtained by analysis, and that these factors can be linked to the next actions. We have successfully proposed a state transition tracking method for state changes of the system data obeying a Poisson distribution. Specifically, this study has dealt with the situation where there are multiple change-points in the Poisson rate parameter, and the most preferred statistical model describing such a situation has been selected using maximum likelihood theory and AIC. Then, the effectiveness of the proposed state transition tracing method has been verified through numerical simulation. By using the proposed state tracing method, the change-points for the Poisson rate parameters can be estimated adequately. It is seen that the information about the change-points is important for identifying the causes that varied the state of systems. If we could detect the causes that varied the state of systems, the next action for the further operation and management can be planed based on the useful information about the causes that varied the state of systems.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number 17K01266: The investigation of data visualization and its application to production and operation management, JSPS KAKENHI Grant Number 18K04611: Evaluation of system performance and reliability under incomplete information environment, and JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2128. We would like to appreciate the grant for our research.

References

1. Samuel, T.R., Pignatiello, J.J., Jr., Calvin, J.A.: Identifying the time of a step change with control charts. Qual. Eng. **10**(3), 521–527 (1998). https://doi.org/10.1080/08982119808919166
2. Samuel, T.R., Pignatiello, J.J., Jr.: Identifying the time of a change in a Poisson rate parameter. Qual. Eng. **10**(4), 637–681 (1998). https://doi.org/10.1080/08982119808919185
3. Perry, M.B., Pignatiello, J.J., Jr.: Estimation of the change point of a normal process mean with a linear trend disturbance in SPC. Qual. Technol. Quant. Manage. **3**(3), 325–334 (2006). https://doi.org/10.1080/16843703.2006.11673118
4. Perry, M.B., Pignatiello, J.J., Jr., Simpson, J.R.: Estimating the change point of a Poisson rate parameter with a linear trend disturbance. Qual. Reliab. Eng. Int. **22**(4), 371–384 (2006). https://doi.org/10.1002/qre.715
5. Takemoto, Y., Arizono, I.: Development and usage of process state transition tracking method for successive data. J. Jpn Ind. Manage. Assoc. **66**(3), 240–248 (2015). https://doi.org/10.11221/jima.66.240

6. Akaike, H.: A new look at the stastical model identification. IEEE Trans. Autom. Control **AC-19**(6), 716–723 (1974). https://doi.org/10.1109/TAC.1974.1100705

Applying Data Driven Approach to Cluster Components for Preventive Maintenance

Ping Yu Hsu, Hong-Tsuen Lei, Ming-Shien Cheng, and Tzu Fan Yuan

Abstract McKone et al. (J Oper Manag 19:39–58, 2001) proposed Total Preventive maintenance (TPM), Just in time (JIT) and Total quality management (TQM) to contribute significantly to manufacturing performance (MP) and TPM could be considered as a part of the manufacturing strategy. The use of preventive maintenance in equipment maintenance could effectively reduce machine occurrence and reduce machine efficiency due to failure (Niu et al. in Reliab Eng Syst Saf 95:786–796, 2010; Panagiotidou and Tagaras in Eur J Oper Res 180:329–353, 2007; Swanson in Int J Prod Econ 70:237–244, 2001). Many studies in the past have applied the concept of total preventive maintenance (TPM) to equipment maintenance to reduce downtime and improve machine efficiency effectively (Panagiotidou and Tagaras in Eur J Oper Res 180:329–353, 2007; Swanson in Int J Prod Econ 70:237–244, 2001). Utilizing preventive maintenance can reduce machine's shutdown and improve the equipment efficiency. The traditional total preventive maintenance methods focused on maintaining single component. The research, however, strives to maintain a group of components to further reduce the maintenance time. The components were clustered into group according to their distributions of lifespans. The clusters that saved the most maintenance costs are recommended to managers for maintenance scheduling. The methodology was applied to an auto component company for experiments. The results showed that OEE was improved from 81 to 84%.

Keywords OEE · TPM · Maintenance strategy · Data driven methodology

---

P. Y. Hsu · H.-T. Lei
Department of Business Administration, National Central University, No.300, Jhongda Rd, Taoyuan County, Jhongli City 32001, Taiwan (R.O.C.)

M.-S. Cheng (✉) · T. F. Yuan
Department of Industrial Engineering and Management, Ming Chi University of Technology, Taishan Dist, No.84, Gongzhuan Rd, New Taipei City 24301, Taiwan (R.O.C.)
e-mail: mscheng@mail.mcut.edu.tw

T. F. Yuan
e-mail: 984401019@cc.ncu.edu.tw

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_19

1 Introduction

McKone et al. [1] proposed Total Preventive Maintenance (TPM), Just in Time (JIT) and Total Quality Management (TQM) to contribute significantly to manufacturing performance (MP) and TPM could be considered as a part of the manufacturing strategy. The use of preventive maintenance in equipment maintenance could effectively reduce machine occurrence and reduce machine efficiency due to failure [2–4].

However, excessively frequent maintenance could also result in wasted time and cost associated with repairing parts. Therefore, in addition to the concept of preventive maintenance in the maintenance process, it was necessary to use the past information to set the appropriate preventive maintenance time. Such a model, in addition to increasing the efficiency of the production line to bring more profits, could also achieve the goal of preventive maintenance at a lower cost [5, 6].

There were some previous studies had considered the issue of reliability-based periodic comprehensive total preventive maintenance (TPM) planning for systems with degraded components [7]. Their goal is to maintain a level of reliability while minimizing overall maintenance related costs. The algorithm has an effective advantage rule that reduces the number of nodes created; the algorithm determines the number of maintenance interventions before each replacement to minimize the total cost over a limited time frame.

The others used different preventive maintenance methods. For example, genetic algorithms are used to calculate TPM plans rather than using manual calculations to minimize maintenance costs [8, 9]. It combined the production planning and the maintenance planning together to reduce downtime on the production line to find the appropriate maintenance time by statistical methods. However, most of these methods were used for preventive maintenance of individual components to minimize prevention.

This study added the concept of total preventive maintenance to routine maintenance to find out when parts should be prevented. In the study, the similarity of the part life data would be analyzed first, and then the similarity of the results would be converted into the distance matrix and the parts would be clustered. Then calculate the time and cost that individual groups could save, and select the appropriate group as the basis for preventive maintenance strategy. In the research work, by analyzing the information of the auto parts manufacturing plant, the appropriate preventive maintenance strategy is proposed. Then, this study putted all the parts on the production line into groups that could be grouped for preventive maintenance, and estimated the time and cost of the preventive maintenance strategy compared to the original maintenance strategy. The results showed that the proposed preventive maintenance strategy can effectively reduce downtime and increase the production rate and OEE of the production line.

However, due to the difference in the maintenance cycle of the parts, the parts would need to be maintained at different times. The more parts there were, the more intensive maintenance was required. When setting the maintenance cycle, consider the maintenance cycle of other parts, adjust the maintenance time appropriately, and maintain the parts with similar maintenance time together. This method was used to improve the prevention and maintenance management of related equipment in a casting foundry plant. This method is used to improve the prevention and maintenance management of related equipment in foundry foundries. The data collection and analysis results for component life have successfully increased OEE from 81 to 84%.

The remainder of this paper is organized as follows. Section 2 describes previous studies. Section 3 establishes the proposed methodology. Section 4 presents the data collection and describes the analysis results. Finally, Sect. 5 presents a discussion and offers a conclusion.

2 Related Work

McKone et al. [1] presented an empirical analysis of Total Production Maintenance (TPM), Just in Time (JIT), Total Quality Management TQM) recognized as a strong contributor to manufacturing performance (MP), and TPM which could be considered as a part of the manufacturing strategy.

The use of preventive maintenance in equipment maintenance could effectively reduce machine occurrence and reduce machine efficiency due to failure [2–4].

The concept of preventive maintenance could be used in routine maintenance, and the time at which parts should be prevented and maintained is different because of the difference in the maintenance cycle of parts. It could be divided into two categories, one is preventive maintenance strategy based on cost orientation and the other is preventive maintenance strategy based on cycle time.

2.1 Preventive Maintenance Strategy Based on Cost Orientation

There were many different preventive maintenance methods in the current study. For example, computing a TPM plan through genetic algorithms minimizes labor and cost [8, 9].

Doostparast et al. [10] proposed a reliability-based preventive maintenance strategy. The goal was to maintain a certain level of reliability at a minimum cost. The proposed cost function strategy includes maintenance costs, replacement costs, system downtime costs, and random failure costs. The reliability of the maintenance equipment was limited by the minimization of the cost function. The cost perspective in reliability theory proposed a preventive maintenance strategy for random working hours.

Chang [11] proposed an optimal maintenance strategy based on equipment usage time, and choose the right maintenance strategy to achieve the lowest average cost rate based on previous strategies [5, 6].

2.2 Preventive Maintenance Strategy Based on Cycle Time

Wan et al. [12] collected a large amount of information generated by the manufacturing industry and set the preventive maintenance based on data driven. The mechanism of preventive maintenance could effectively improve the efficiency of manufacturing production. Cassady and Kutanoglu [13], Sortrakul et al. [14] investigated the preventive maintenance program in conjunction with the production plan. Because of the production plan and maintenance plan would affect each other.

Cassady and Kutanoglu [13] combined production plans to evaluate production plans and maintenance plans would affect each other. To consider the point in time when the two projects are executed to allocate the two projects more efficiently and minimize the time required to maintain the plan.

In the past, some of the relevant research focused on time points and costs to make appropriate preventive maintenance strategies and partial researches had focused on methodological research. However, the difference in this study was that the point of time consideration was to analyze the similarity between the life data of the parts to determine the time point of group prevention and maintenance. This investigation proposed to group parts and to replace parts with similar service life by data driven method. The machine could be maintained in accordance with the least costly preventive maintenance strategy to effectively improve production efficiency.

However, due to the difference in the maintenance cycle of the parts, the parts would need to be maintained at different times. The more parts there were, the more intensive maintenance was required. When setting the maintenance cycle, consider the maintenance cycle of other parts, adjust the maintenance time appropriately, and maintain the parts with similar maintenance time together.

3 Methodology

In this study, the data similarity analysis is first carried out, and the collected life data of each part is judged by Kolmogorov–Smirnov statistical test method. Then group the parts and list all possible combinations according to the number of parts (Figs. 1and 2).

The results of the Kolmogorov–Smirnov statistical test method for the two parts were recorded in a matrix. The part with p-value < 0.05 was considered as the correlation between the two parts was not significant, and the remaining part was based on the detection result p-value of two identical parts. Calculate the difference

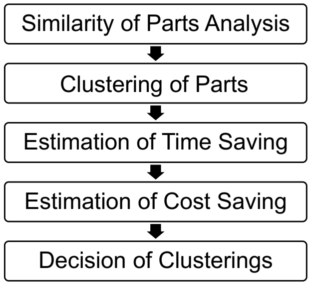


Fig. 1 Frame of investigation

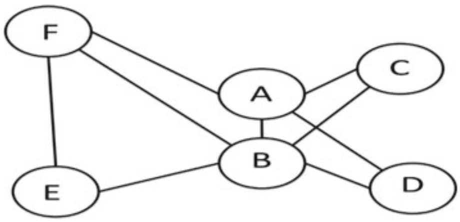


Fig. 2 Distance of two pairs of related two parts

between each piece of data in the matrix and 1 as a basis for measuring the similarity between parts.

For example, if 7 parts are used for preventive maintenance, they are represented by A, B, C, D, E, F, and G respectively. Before performing the test, it is necessary to first create a blank matrix of the corresponding size according to the number of parts, and perform all parts separately for ks-test. The p-value of the final result is recorded in the matrix before the start of the experiment. If the p-value < 0.05 indicates that the result is not significant, there is no degree of similarity between the two parts. In this study, it could be divided into a group of parts that are jointly maintained, representing the fault cycle data of the parts with similarity. These groups could be considered as groups that might be combined for preventive maintenance. And so on, to find two pairs of related parts. Figure 3 was a result of the grouping of two parts. And then to search for groups with a group of three, etc.

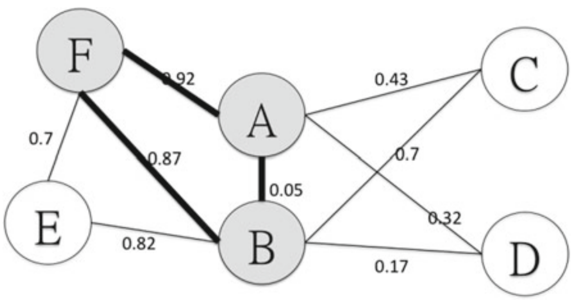


Fig. 3 Distance of two pairs of related three parts

After grouping parts, carry out preventive maintenance strategies for machine maintenance. It can save downtime for individual parts due to failure, but it takes one maintenance time to perform machine maintenance.

According to the preventive maintenance strategy, the benefits of reducing the stoppage time can be saved. But what needs to be paid is the waste of parts replacement caused by advance maintenance. The difference in the middle is the profit that can be obtained according to the preventive maintenance strategy.

4 Data Collection and Analysis

The data collected in this study are the production materials, order materials and power usage information of the foundry division of a major auto parts manufacturer located in Taoyuan, Taiwan from July 01 2016 to June 30 2017. This auto parts car manufacturer is the first tier supplier of Toyota, Nissan, Ford, and other car manufacturers. It mainly produces aluminum alloy wheels, brake parts, engine parts, and so on. The casting process is the first manufacturing process for alloy wheels, brake parts and engine parts.

According to the results of grouping, preventive maintenance should first find out the time points between the groups covering 95% of the data. Select the shortest time in the group as the basis for preventive maintenance. Although it will incur additional costs, it can improve the reliability of machine operation and reduce the chance of parts failure.

The estimation method could be used to calculate the downtime and cost required for a long time in accordance with the preventive maintenance strategy. To compare the calculated results with the data currently collected proves that preventive maintenance can not only effectively reduce downtime but also achieve the goal at the lowest cost. In the case, all parts of the production line are grouped. A total of 70 parts are divided into the following 13 groups according to the similarity and cost considerations of the life data of the parts. The results of one-time preventive maintenance are shown as Tables 1 and 2.

Table 1 Results of one-time preventive maintenance

Item	Clustering	Cost of maintenance	Time of maintenance
1	Mold base, lower top cylinder, upper top cylinder, lower mold cylinder	17,010.7087	204.54421
2	Upper mold cylinder, solenoid valve	7417.8479	176.25806
3	No. 2/No.1 vibrating machine	4676.254	155.08571
4	No. 1/No. 2 Skirt, No. 3# vibrating machine	8160.6563	146.29152
5	108 Vibrating machine, Iron sand machine	5905.152	144.74231
6	Up bloom room, rotary positioning cylinder, Rotary cylinder	13,033.8172	69.75
7	Clamping cylinder, push cylinder	7420.1727	38.5
8	Sandbox, oil pressure unit	3303.969	57.5
9	Rotary cylinder, photoelectric unit	7499.8267	21.2
10	Up bloom room, low bloom room 、rotary positioning cylinder	13,561.7622	37.65
11	Extrusion cylinder, fork rod cylinder	6649.3667	27
12	RCS, No. 1 Host、magnetic separator	13,539.545	88.0817
13	No. 3 Hoist, No. 13 Belt	3001.7	46
14	No. 2 Hoist, RCT	4468.096	74.4

Table 2 Comparison of maintenance time before and after preventive maintenance

Item	Clustering	Time of Maintenance (before)	Time of Maintenance (after)	Improvement (%)
1	Mold base, lower top cylinder, upper top cylinder, lower mold cylinder	260	204.54421	21
2	Upper mold cylinder, solenoid valve	214	176.25806	18
3	No. 2/No.1 Vibrating machine	215	155.08571	28
4	No. 1/No. 2 Skirt, No. 3# vibrating machine	188	146.29152	22
5	108 Vibrating machine, Iron sand machine	196	144.74231	26

(continued)

Table 2 (continued)

Item	Clustering	Time of Maintenance (before)	Time of Maintenance (after)	Improvement (%)
6	Up bloom room, rotary positioning cylinder, rotary cylinder	122	69.75	43
7	Clamping cylinder, push cylinder	79	38.5	51
8	Sandbox, oil pressure unit	128	57.5	55
9	Rotary cylinder, photoelectric unit	60	21.2	65
10	Up Bloom Room, low Bloom room, rotary positioning cylinder	69	37.65	45
11	Extrusion cylinder, fork rod cylinder	63	27	57
12	RCS, No. 1Host, magnetic separator	203	88.0817	56
13	No.3 Hoist, No. 13 Belt	127	46	64
14	No. 2 Hoist, RCT	118	74.4	37

5 Results and Future Study

This study analyzed the similarity between the life data of the parts to determine the time point of group prevention and maintenance. The machine can be maintained in accordance with the least costly preventive maintenance strategy to effectively improve production efficiency.

The theoretical contribution of this paper contribution of this paper is threefold:

1. The results showed that OEE was improved from 81 to 84%.
2. The working hours of the project are unchanged and the part of the directly usable working hours is increased from the original 484,072 min to 501,606 min.
3. In the case of the study, the parts were divided into 14 groups, and according to the obvious effect in saving cost and saving time.

In the study, the production line of high temperature and high heat is analyzed. The structure and operation of the machine are relatively simple, the abnormal condition of the production line is frequent, but the repair time is relatively short. In future research, this analytical method can be applied to equipment in different fields for analysis and research, such as application in high-tech industries, with more precise processes and more expensive equipment, and electrical dipping industry, etc. It is applied to the textile industry with simple equipment construction.

It is hoped that the equipment in different fields will be proved, so that the preventive maintenance methods in research can be more widely used.

References

1. McKone, K.E., Schroeder, R.G., Cua, K.O.: The impact of total productive maintenance practices on manufacturing performance. J. Oper. Manag. **19**(1), 39–58 (2001)
2. Niu, G., Yang, B.S., Pecht, M.: Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance. Reliab. Eng. Syst. Saf. **95**(7), 786–796 (2010)
3. Panagiotidou, S., Tagaras, G.: Optimal preventive maintenance for equipment with two quality states and general failure time distributions. Eur. J. Oper. Res. **180**(1), 329–353 (2007)
4. Swanson, L.: Linking maintenance strategies to performance. Int. J. Prod. Econ. **70**(3), 237–244 (2001)
5. Budai, G., Huisman, D., Dekker, R.: Scheduling preventive railway maintenance activities. J. Oper. Res. Soc. **57**(9), 1035–1044 (2006)
6. Eti, M.C., Ogaji, S.O.T., Probert, S.D.: Reducing the cost of preventive maintenance (PM) through adopting a proactive reliability-focused culture. Appl. Energy **83**(11), 1235–1248 (2006)
7. Tsai, Y.T., Wang, K.S., Teng, H.Y.: Optimizing preventive maintenance for mechanical components using genetic algorithms. Reliab. Eng. Syst. Saf. **74**(1), 89–97 (2001)
8. Chan, F.T., Chung, S.H., Chan, L.Y., Finke, G., Tiwari, M.K.: Solving distributed FMS scheduling problems subject to maintenance: genetic algorithms approach. Robot. Comput. Integr. Manuf. **22**(5–6), 493–504 (2006)
9. Saranga, H., Kumar, U.D.: Optimization of aircraft maintenance/support infrastructure using genetic algorithms—level of repair analysis. Ann. Oper. Res. **143**(1), 91 (2006)
10. Doostparast, M., Kolahan, F., Doostparast, M.: A reliability-based approach to optimize preventive maintenance scheduling for coherent systems. Reliab. Eng. Syst. Saf. **126**, 98–106 (2014)
11. Chang, C.C.: Optimum preventive maintenance policies for systems subject to random working times, replacement, and minimal repair. Comput. Ind. Eng. **67**, 185–194 (2014)
12. Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., Vasilakos, A.V.: A manufacturing big data solution for active preventive maintenance. IEEE Trans. Ind. Inf. **13**(4), 2039–2047 (2017)
13. Cassady, C.R., Kutanoglu, E.: Integrating preventive maintenance planning and production scheduling for a single machine. IEEE Trans. Reliab. **54**(2), 304–309 (2005)
14. Sortrakul, N., Nachtmann, H.L., Cassady, C.R.: Genetic algorithms for integrated preventive maintenance planning and production scheduling for a single machine. Comput. Ind. **56**(2), 161–168 (2005)

Process Simulation of Compression Molding Process and Effect of Fiber Content on Recycled Polymer Natural Fiber Composites Using Moldflow Analysis

Vardaan Chauhan, Timo Kärki, and Juha Varis

Abstract This paper presents the process simulation and flow of material in the cavity of the mold during the compression molding process. The composite material used as the charge in the molding process is comprised of recycled polymers waste extracted from various automotive parts and reinforced with engineered wood fibers. The compression molding tool is a re-design version of an automotive battery cover that is currently in application in various car models. The newly designed component is suitable to be manufactured using recycled polymer composite material. Some additives were added to the composite material to make it more viscous and to aid in the flow of material in the mold. The simulation models are built-in Moldflow with the same part geometry and processing conditions were kept the same for all the composite blends. Later, all the composite blend models were compared with only recycled polymer models to see the effects of fiber material on the output of the compression molding process. The final simulation shows that the 10% fiber content in the composite material exhibits the most promising option with no voids or cavities and less fill time of the entire part geometry.

Keywords Process simulation · Recycled polymers · Natural fibers · Compression molding · Moldflow

1 Introduction

Natural fiber polymer composites (NFPCs) have been widely used in the automobile sector as a substitute material for metal and plastics parts such as door panels, seat backrests, door linings, and luggage compartments [1–3]. Fibers such as wood, flax hemp, sisal, and kenaf are commonly used by car manufacturers like Daimler and BMW in manufacturing automotive components [3]. There are different manufacturing methods used to produced NFPCs components. These include compression molding [4, 5], injection molding [4–6], resin transfer molding [7], and extrusion.

V. Chauhan (✉) · T. Kärki · J. Varis
LUT University, Lappeenranta, Finland
e-mail: vardaan.chauhan@lut.fi

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_20

However, compression molding is a popular process because it is a high-volume production process and is capable of processing even a high volume of fiber content at high temperatures and pressure [8]. Comparatively injection molding is also a high-volume production process but it difficult to process high fiber content because the fiber reduces the mobility of material in the mixing barrel and can also block the orifice of the machine connected to the mold. Though compression molding process also possesses its limitation mainly when it comes to processing direction since it can operate only operate on the z-axis. This in turn leads to certain design limitations or having a complex compression molding tool, which increases the overall tooling cost.

Therefore, a tool such as Moldflow can aid in simulating the manufacturing process virtually thereby investigating whether the component design is optimal for NFPCs material and even suggest the best material composition for that design. Moldflow has been used in previous researches especially for injection molding and virgin polymers, where researchers were able to identify optimal gate location, processing parameters and reduce cycle time for manufacturing a particular component [9, 10]. However, it can also be used to simulate the flow and calculate the fill time of the compression molding process for NFPCs material. This can help in reducing the overall development time involved when using these recycled polymers alongside natural fibers since Moldflow can reduce trial and error methods during the manufacturing process and can provide accurate data about optimal processing parameters and the quality of the final product.

The study aims to examine the flow of NFPC material in the compression molding process. In the study, varying percentages of engineered wood were mixed with recycled polymer extracted from automotive polymer waste, and their flow was analyzed, and the optimum ratio of the wood-to-recycled polymer was documented for the composite material.

2 Material and Methods

2.1 Test Model

The test model for this experiment was a car battery cover, which was designed in-house using Solidworks 2017 and later imported into Moldflow software for flow analysis purposes. The part was designed to replace the existing battery cover design that is made from virgin polymer using the injection molding process. The designed part is suitable for the compression molding process and NFPC material. The test model is meshed using 3D mesh in Moldflow and is shown in Fig. 1.

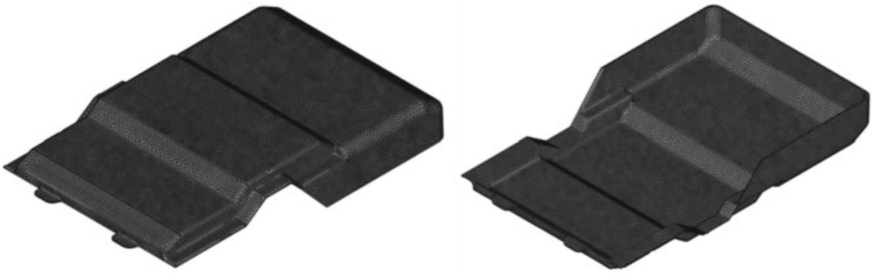


Fig. 1 Meshed test model using 3D mesh in Moldflow

2.2 Material Composition

The polymer wastes used in the experiment and simulation were extracted from automobile, motorbike, and truck parts, such as frontal fairings, truck bumpers, and door panels as shown in Fig. 2. The waste is a mixture of primarily ABS and other styrene-based polymers such as polystyrene, styrene-acrylonitrile resin, and acrylonitrile styrene acrylate. The waste polymers were crushed in a granulator before material testing. The material properties such as density, melt flow index, and mechanical performance of recycled polymer waste were obtained in the laboratory by testing and few missing values were taken from a generic ABS material database present in Moldflow, later the entire information was added to the Moldflow material database.

The material properties of engineering wood were already available in the Moldflow material database. Additionally, material data of lubricant and elastomer used for easy removal of the part after processing and improve the flowability of the material in the mold is also added to the Moldflow database. This is done because the recycled polymer waste might contain some impurities which can disrupt the flow of material. The composite composition is shown in Table 1.

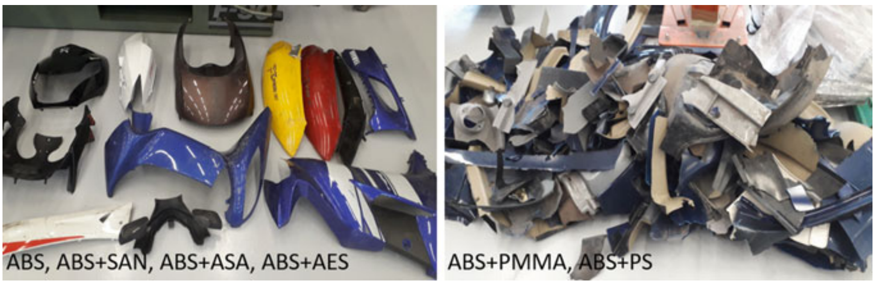


Fig. 2 Waste polymers used in experiment and simulation

Table 1 Natural fiber polymer composite material composition

Blend	Polymer (%)	Fiber (%)	Lubricant (%)	Elastomer (%)
Blend 1	94	–	3	3
Blend 2	84	10	3	3
Blend 3	74	20	3	3
Blend 4	64	30	3	3

Table 2 Processing parameters used in Moldflow simulation

Mold temperature	Melt temperature	Pressure	Compression time
50 °C	230 °C	40 ton	60 s

2.3 Processing Parameters

The processing parameters input into the Moldflow simulation were kept constant for all the composite blends to see the effect of the addition of fiber on the flowability of polymer material in the mold. The processing parameters used in the experiment are shown in Table 2.

3 Result and Discussion

The simulation results of all the blends are shown in Fig. 3. From the figure, it can be observed that for blends 1, 2 and 3 the part quality was good and acceptable. However, for blend 4 the part quality is not desirable. Therefore, up to 20% of engineering wood fibers in the composite are acceptable while any further increase could lead to the part not filled. This is mainly due to high fiber content which greatly affects the flowability of the composite material inside the mold even with the presence of 3% elastomers inside. For both 10% and 20% fiber composite the part quality was good, but it is visible that in 10% fiber composite the flow is much better as it has less resistance from the fibers.

The fill time for blend 1 was the lowest at 1.40 s this is because the blend consists of only polymer material and no fiber contents which could have reduced the flowability of material in the mold and thus could have to increase the filling time. For blends 2 and 3, the filling was about 2.27 s and 3.85 s respectively. This indicates the effect of increasing fiber content in the composite material on the filling time and this could affect the part quality also. Finally, blend 4 has the highest filling time at 9.65 s probably due to the presence of high fiber content. Since the part from blend 4 did not fill even though the compression molding pressing time was 60 s, this indicates that the material cooled down before it can fill the mold.

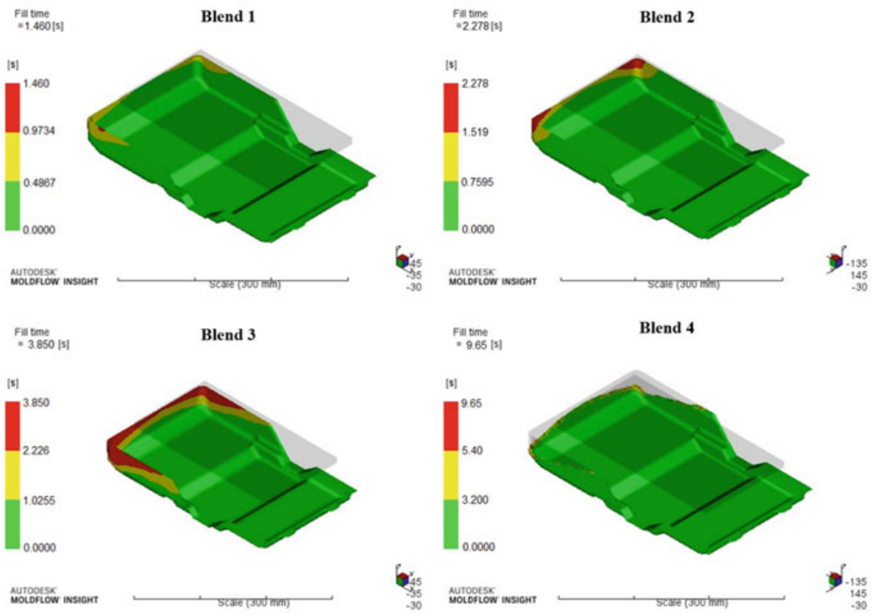


Fig. 3 Moldflow simulation results for all the blends

Based on the simulation, blend 2 with 10% engineered wood fiber is the best solution amongst the three NFPCs material when using this recycled polymer waste from automobiles. Blend 2 with 20% engineered wood fiber is also an option but due to slightly higher filling time and the possibility of some variations between simulation and production of part via a compression molding process, it was not selected for the final part production stage. The result from compression molding of blend 2 is shown in Fig. 4.

Fig. 4 Sample produced by compression molding of blend 2

4 Conclusion

In this study, the process simulation and NFPCs material flow in the mold of compression molding process were performed using Moldflow software. Moldflow offers high reliability in process simulation and can greatly save testing costs and time. The optimum wood fiber content for NFPCs material consisting of recycled polymer waste was found to be 10–20% fiber content. Further increasing the fiber content leads to poor quality of the molded part. Based on the filling time, 10% fiber content composite exhibits the lowest value and thus meant lower cycle time during production. Later, based on Moldflow analysis, one sample from 10% fiber content was produced in the lab and the sample came out was in desirable condition.

References

1. Chauhan, V., Kärki, T., Varis, J.: Review of natural fiber-reinforced engineering plastic composites, their applications in the transportation sector and processing techniques. J. Thermoplast Compos. Mater. (2019). https://doi.org/10.1177/0892705719889095
2. Holbery, J., Houston, D.: Natural-fiber-reinforced polymer composites in automotive applications. J. Miner. Met. Mater. Soc. **58**(11), 80–86 (2006). https://doi.org/10.1007/s11837-006-0234-2
3. Zhao, D., Zhou, Z.: Applications of lightweight composites in automotive industries. Am. Chem. Soc. 143–158 (2014). https://doi.org/10.1021/bk-2014-1175.ch009
4. Bledzki, A.K., Faruk, O.: Wood fiber reinforced polypropylene composites: compression and injection Molding Process. Polym. Plast. Technol. Eng. **43**(3), 871–888 (2004). https://doi.org/10.1081/PPT-120038068
5. Takemura, K., Minekage, Y.: Effect of molding condition on tensile properties of hemp fiber reinforced composite. Adv. Compos. Mater. **16**(4), 385–394 (2007). https://doi.org/10.1163/156855107782325177
6. Liu, W., Drzal, L.T., Mohanty, A.K., Misra, M.: Influence of processing methods and fiber length on physical properties of kenaf fiber reinforced soy based biocomposites. Compos. B Eng. **38**(3), 352–359 (2007). https://doi.org/10.1016/j.compositesb.2006.05.003
7. Lokesh, P., Surya Kumari, T.S.A., Gopi, R., Loganathan, G.B.: A study on mechanical properties of bamboo fiber reinforced polymer composite. Mater. Today Proc. **22**(3), 897–903 (2020). https://doi.org/10.1016/j.matpr.2019.11.100
8. Hasan, K.M.F., Horváth, P.G., Kóczán, Z., Alpár, T.: Thermo-mechanical properties of pretreated coir fiber and fibrous chips reinforced multilayered. Sci. Rep. **11**(1), 3618 (2021). https://doi.org/10.1038/s41598-021-83140-0
9. Salunke, M., Kate, R., Lomate, V., Sopal, G.: Injection molding methods design, optimization, simulation of plastic toy building block by mold flow analysis. Int. J. Mech. Eng. Technol. **6**(6), 33–42 (2016)
10. Vishnuvarthanan, M., Panda, R., Ilangovan, S.: Optimization of injection Molding cycle time using Moldflow analysis. Middle East J. Sci. Res. **13**(7), 944–946 (2013). https://doi.org/10.5829/idosi.mejsr.2013.13.7.2934

A Combined Scheduling and Simulation Method to Analyze the Performance of the Dual-Robot In-Line Stocker

Jaewoo Chung

Abstract The in-line stocker is a new type of the automated material handling system that has begun to be used in the display industry. This not only moves unit loads under processing across different locations within manufacturing facilities, similar to the automated guided vehicle (AGV), but also stores them on its shelves, similar to the automated storage and retrieval system (ASRS), which can significantly reduce space for material handling. However, traffic rates are generally very high inside the in-line stocker and two robots serve along the single lane, which is the dual-robot in-line stocker (DRIS). One difficulty in applying the DRIS to shop floors is that the exact transport capacity of the unit DRIS is not known. This paper develops an analytical model to estimate the capacity of the DRIS based on a combined scheduling and simulation method. It calculates movements of two robots over time in the space consisting of time and location and precisely measure **necessary time for waiting or backtracking to avoid collision of two robots.** An experimental analysis was conducted to validate the correctness and usefulness of the model based on the data used in an actual manufacturing site. The analysis result illustrates that the average processing capacity of the DRIS increases compared to the SRIS (single robot in-line stocker) as the length of the DIRS increases, which is consistent with the expectation of practitioners in the industry. The paper also verifies that it is necessary to carefully determine the operating specifications in actual uses since the transport capacity of the DIRS varies considerably by its operating parameters, which can be optimized by the analytical model.

Keywords In-line stocker · Dual-robot AS//RS · Automated material handling system · Scheduling and simulation · Simulation modeling

J. Chung (✉)
Department of Operational Management, School of Business Administration, Kyungpook National University, Daegu, Republic of Korea
e-mail: chung@knu.ac.kr

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_21

1 Introduction

The in-line stocker is a new type of the automated material handling system (AMHS), which carries out storage and transportation of unit loads under processing in the manufacturing or warehousing facilities. It was the display industry that this type of system first came into use and it is being adopted by other industries such as semiconductor or battery industries. The display industry conventionally used the AGV (automated guided vehicle) system to move unit loads and the automated storage and retrieval system (ASRS) were used separately to store unit loads between processing machines in fabrication facilities. However, as the sizes of the display products increased to meet market demands, proportionally more spaces were required for material handling and consequently the higher investment cost was required to build a new fabrication facility. For reducing arising cost pressure, engineers in the industry devised to use the ASRS for not only storage but also transportation as illustrated in Fig. 1. A crane robot is moving along the lane in the center of the in-line stocker the same as the ASRS for material handing as explained in (a) of Fig. 1. Processing machines are connected in the ground level of the in-line stocker and 4–6 stories of the shelves are placed in the upper level in (b) of Fig. 1.

By using this design, the display industry typically uses tens of the in-line stockers in a fabrication facility and could save the space cost of the clean room significantly; however, the transportation rate of the unit in-line stocker became much higher compared to the conventional ASRS. To process the higher transport rate, the engineers adopted a dual-robot in-line stocker (DRIS) that two robots are serving along the lane in the unit in-line stocker as seen in (a) of Fig. 1. Due to this structure, one robot has to wait the other when their paths are overlapped each other, which is called blocking. In an even worse case, a robot sometimes has to move to the opposite direction of an original travel to avoid collision, which is called *back tracking*. In addition, the travel boundary of one robot is limited in the lane due to the other robot. Several studies related to the dual-robot in-line stocker have been found in previous literature, but the names of the system are slightly different. Some named it as the inline stalker [1, 3 and 4] dual stocker robot [5], and dual robot stalker system [1]. In this work, we name this system as the DRIS to make it more general.

One difficulty in applying the DRIS to shop floors is that the exact transport capacity of the unit DRIS is not known. One can easily see that the upper bound of

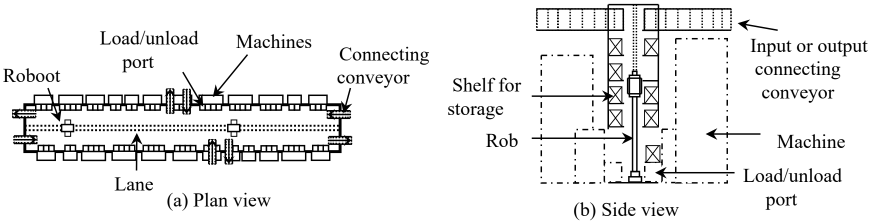


Fig. 1 Structure of the In-line stocker

its transport capacity is lower than two times that of the single robot in-line stocker (SRIS) due to blocking or back tracking. In practice, a very rough approximation is used to estimate the transport capacity of the DRIS, which is 1.5 or 1.6 times that of the SRIS. Note that there are reasonable methods to estimate the transport capacity of a SRIS [2, 7]. Such rough estimates on a DRIS lead to design losses such as an excessive investment on a DRIS even when a SRIS is sufficient, or long delay times due to not being able to handle the transport jobs during a peak time even with two robots in a DRIS.

One of the difficulties estimating the transport capacity of a DRIS comes from the fact that it depends on several input parameters such as its length of the lane, the travel zones of the two robots [4, 8], and origins and destinations of transport jobs. This paper develops an analytical model to estimate the transport capacity of the DRIS under various conditions based on a combined scheduling and simulation approach. The new method expresses the movements of the two robots in a two dimensional space, which consist of time and location, and makes detailed movement schedules of given transport jobs on the two robots in the space.

2 Model Development

2.1 Symbols

This paper uses the following symbols to develop the analytical model. There are two robots a and b, and moving boundaries of the two robots are divided into three zones based on which robot or robots can travel into the zone. Robot a only can move in Zone A, Robot b only can move in Zone C, and both the robots can move in Zone B. The home position of Robot a is the left end of the lane, and that of Robot b is the right end of the lane.

N: Total number of jobs to be transported, $k \in \{1, 2, \ldots, N\}$.
j: Index of a robot, $j \in \{a, b\}$.
k_j: Index of a job assigned to robot j.
l: Lane length of the DRIS.
l^j_{min}: the minimum location that robot j can travel, $0 \le l^j_{max} \le l$.
l^j_{max}: the maximum location that robot j can travel, $0 \le l^j_{max} \le l$.
o_{jk}: Origin of job k that is assigned to robot j, $k \in \{1, 2, \ldots, N\}$.
d_{jk}: Destination of job k that is assigned to robot j, $k \in \{1, 2, \ldots, N\}$.
c^t_{jk}: Location of robot j processing job k at time t, if j = a, $0 \le c^t_{jk} \le l^j_{max}$, and if j = b, $l^j_{min} \le c^t_{jk} \le l$.
c_{jk}: Location of robot j right after completing job k.
m^{λ}_{jk}: Travel distance of robot j to load job k.
m^{μ}_{jk}: Travel distance of robot j to unload job k.
e^{λ}_{jk}: Travel time of robot j to load job k.

e^{μ}_{jk}: Travel time of robot j to unload job k.

w_{jk}: Travel time plus waiting or backtracking time of robot j to avoid a collision while transporting job k.

δ^{t}_{jk}: Moving status of robot j at time t while transporting job k. $\delta^{t}_{jk} = 0$, if the robot stops, $\delta^{t}_{jk} = -1$, if it is moving from the right to the left, and $\delta^{t}_{jk} = 1$, if it is moving from the left to the right.

I^{t}_{jk}: Binary variable indicating the status of robot j at time t while transporting job k. $\mathrm{I}^{\mathrm{t}}_{\mathrm{jk}} = 0$, if the robot stops or travelling for avoidance, and $\delta^{\mathrm{t}}_{\mathrm{jk}} = 1$, if it is moving.

u: Loading and unloading time. It is assumed that loading and unloading times are the same.

v: Moving time of a robot (e.g., meter per minute).

T_{jk}: Total time to move job k by robot j.

s_j: Earliest start (or available) time of robot j to move the next job.

Δ: Minimum clearing distance to avoid the collision of two robots.

2.2 Generation of Transport Jobs

The capacity of a DRIS with various input parameters is calculated through a simulation procedure and the generation of the transport jobs for the procedure consists of two steps. First, the origins and destinations of all N jobs are generated separately. Each job has a origin and destination within length l. These jobs will be moved independently if a SRIS is used, but if a DRIS is used, some of these jobs should be transported by the two robots. If the origin and destination of a job are in Zones A and C, or C and A respectively, the job cannot be completed by one robot, therefore, the job must be broken into two disjointed jobs. For an example, if the origin and destination of a transport job are in zones A and C, Robot a moves the job from Zone A to Zone B, and then Robot b moves it to its final destination in Zone C from Zone B. For convenience of explanation, let us call this type of transport jobs as a long distance job. If there is a total of α long distance jobs out of N jobs created in the first step, the second step of the procedure generates α stopover points in Zone B to break the long distance jobs into two jobs. After the second step, a total of $\mathrm{N} + \alpha$ will be generated for simulation. We can easily imagine that if α is greater, then the capacity of a DRIS would be drop since more jobs should be disjointed into two jobs.

2.3 Scheduling Jobs

Scheduling transport jobs starts by selecting a robot out of the two based on their earliest start times (EST). The one that has the smallest EST is selected and if the EST's of the two are the same, an arbitrary one can be selected. Once a robot is selected then a transport job is selected out of the jobs that are able to be moved by the selected robot based on the FCFS (first-come first-service). If both the robot and job are selected, then its scheduling is done by calculating all the job processing variables defined in Sect. 2.1. A robot performs four steps of activities to complete a job, which are 'travel to load', 'loading', 'travel to unload', and 'unloading'. This procedure is repeatedly done one job at a time until all the jobs of N+α are completed and various statistics such as the total completed time, blocking and backtracking rate of robots, and transport capacity are computed in the end.

Figure 2 explains how the scheduling procedure internally deals with transport jobs to consider locations of two robots over time so that they do not collide with each other, which is named as the time-location graph by this paper. The figure consists of two dimensional space, which are time in x−axis and locations of two robots in y−axis. The total length of the DRIS, which is l, is in the top of y−axis. Three zones explained above are also marked in the y−axis with $l_{j=a}^{\max}$ and $l_{j=b}^{\min}$. There are all four transport jobs in the figure as an illustrative example, which are Jobs m − 1 and m to be moved by Robot a, and Jobs n − 1 and n to be moved by Robot b. Note that neither waiting nor back tracking is required to move these four jobs by Robots a and b. The locations of Robot a over time is in the solid line while that of Robot b is in the broken line in the figure. Job n − 1 assigned to Robot b is completed at time t_{n-1} and Job m − 1 assigned to Robot a is completed at time t_{m-1}. For the second jobs of each robot are completed at times $s_{j=a}$ and $s_{j=b}$ respectively. For scheduling the next job, Robot b will be selected instead of Robot a because its EST is smaller.

Using the time-location graph, the scheduling procedure handles waiting and backtracking whenever necessary. Assume that the origin of the next job to Robot b, which is $o_{j=b,n+1}$, is in between $d_{j=a,k=m}$ and $l_{j=b}^{\min}$. Then, Robot b will move to $d_{j=a,k=m} + \Delta$ immediately after completing Job n but has to wait there until time $s_{j=a}$ for avoiding a collision since Robot a is unloading job at $d_{j=a,k=m}$. After Robot a completes unloading Job m at $d_{j=a,k=m}$, it must backtracks to the direction of its home position, which is 0, even though the origin of its next job is in the greater position than its current position, which is $d_{j=a,k=m}$ to avoid a collision with Robot b

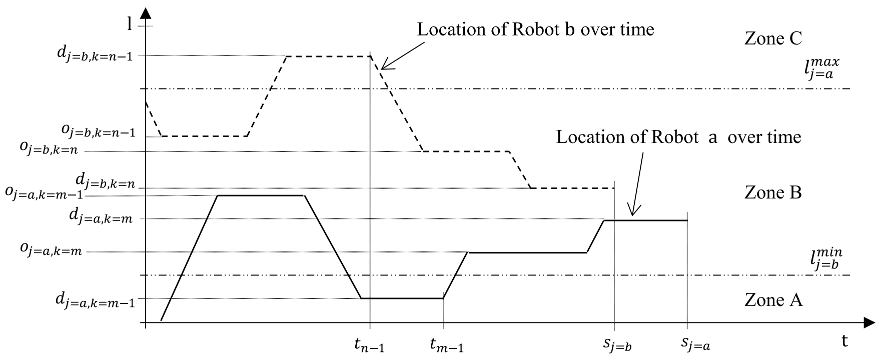


Fig. 2 Time-location graph for jobs and assigned to Robot, and jobs and assigned to Robot

that is already under the way to load Job n + 1 at $o_{j=b,k=n+1}$. Robot a has to backtrack all the way to $o_{j=b,k=n+1} - \Delta$.

All the variables defined in Sect. 2.1 are calculated during transport jobs for all activities as explained blow. Formula (1) computes the current location over time during transporting activities are performed. Formulas (2) and (4) compute travel distances and corresponding transport times elapsed are calculated through formulas (3) and (5). Formulas (6)–(9) update necessary variables after a transport job is completed.

$$c_{jk}^{t} = c_{jk}^{t-1} + \delta_{jk}^{t} \tag{1}$$

$$m_{jk}^{\lambda} = \left| o_{jk} - c_{j,k-1} \right| \tag{2}$$

$$e_{jk}^{\lambda} = v \cdot m_{jk}^{\lambda} \tag{3}$$

$$m_{jk}^{\mu} = \left| o_{jk} - d_{j,k} \right| \tag{4}$$

$$e_{jk}^{\mu} = v \cdot m_{jk}^{\mu} \tag{5}$$

$$w_{jk} = \sum_{t=T_{j,k-1}}^{T_{jk}} I_{jk}^{t} \tag{6}$$

$$T_{jk} = e_{jk}^{\lambda} + e_{jk}^{\mu} + 2\mathrm{u} + w_{jk} \tag{7}$$

$$c_{jk} = c_{j,k-1} + v \sum_{h=1}^{T_{jk}} \delta_{jk}^{c_{j,k-1}+1} \tag{8}$$

$$s_{j} = \max_{jk}\left(c_{jk}\right) = \sum_{h=1}^{k} T_{jh} \tag{9}$$

3 Experimental Analysis

3.1 Input Data

The performance of the DRIS is compared with that of the SRIS (single robot in-line stocker) under various conditions based on the analysis model developed in Sect. 2. The analysis model of the SRIS is similar to that of the SRIS but much simpler because neither waiting nor backing movement is necessary. The input data used for this analysis was collected from a display firm in Republic of Korea. The travel speed of the robot (v) is 60m/min, time for pick-up and drop-off (u) is 20s, and minimum clearing distance (Δ) is 2m to avoid the collision. The analysis changes two types of input data including the length of the in-line stocker and the lengths of Zones A and C. As discussed above, it is a general idea that if the lengths of Zones A and C are longer, more travel jobs should be disjointed; hence, the performance of the DRIS would be lower. But the practitioners want to have precise measurements. The length of the in-line stocker (l) is changed from 5–60m. And the lengths of Zones A and C are the same and are changed from 2–26 m in the analysis. The total number of transport jobs (N) created for each simulation run is 1000 and each simulation run is repeated 1000 times for obtaining the results in a steady state. The analysis observes various output variables including the average completed time of all N jobs and the average waiting or backtracking time ratios of the DRIS in various conditions.

3.2 Analysis Results

Figure 3 compares the average times to complete all the transport jobs (N) with the single and dual in-line stockers. In this analysis, the length of Zones A and B was fixed at two meters for any length of the in-line stocker. The average completed time steadily increases as the length of the SRIS is greater, which is the broken line in the figure. However, the average completed time of the DRIS initially decreases as its length is greater and it again increases after its length reaches to 9 m, which is depicted in the solid line of the figure. The reason why the total completed time decreases is because more jobs are disjointed because the length ratio of Zones A and C to Zone B is high when the length of the DRIS is smaller than 9 m. In sum, when the length of Zones A and B is 2 m and the length of Zone B is 5 m, the total completed time is minimized with the DRIS of the length 9 m, of which the ratio is 4/9.

The second row of Table 1 below shows how the ratio of the disjointed jobs is decreases as the length of the DRIS increases in the first row. The length of Zones A and C is also 2 m in this analysis. When the length of Zones A and C is 2 m for each and the length of Zone B is only one meter, the ratio of the disjointed jobs is as high as 1.28 when its total length is 5 m, which means that 28% out of the total N jobs are disjointed. One can observe that the disjointed ratio decreases as the length increases and further it initially quickly decreases and it is saturated to 1.01 after its length reaches at 25 m.

The third row of Table 1 gives the waiting or backtracking time ratio of the DRIS as its length increases in the first row. Note that the clearing distance to avoid collision is 2 m. When the length of the DRIS is only 5 m, an 80% of the robots should wait or backtrack out of the total time to complete all transport jobs. In this case, the capacity ratio of the DRIS to the SRIS is 0.96, which means that the capacity of the DRIS is lower than that of the SRIS as described in the fourth row of the table. The waiting

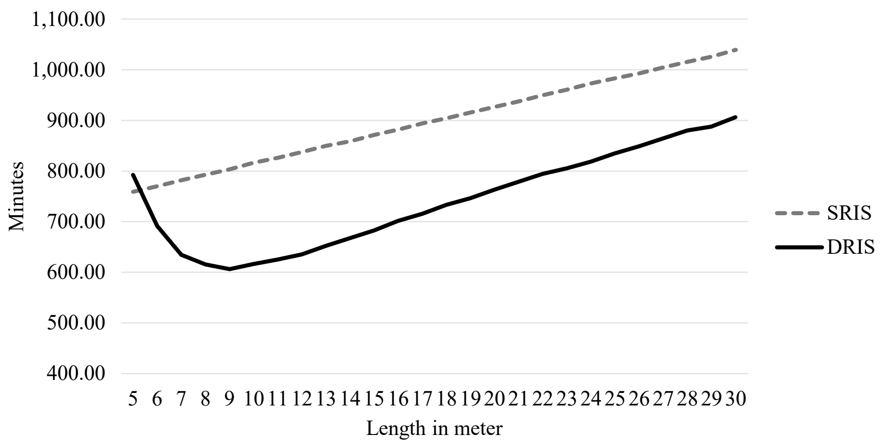


Fig. 3 Comparison of the average completed times to transport all jobs

or backtracking time ratio is minimized at the length of 9 m and the capacity of the DRIS is also maximized at this point with the ratio of 1.33. The ratio of the waiting and backtracking time increases again from the length of 11 m because the length of Zones A and C is fixed to 2 m while the total length keeps increasing. It is expected that this waiting and backtracking ratio can be decreased by increasing the length of Zones A and C as the length of the DRIS increase. The highest capacity ratio of the DRIS compared to the SRIS is 1.33 at its length of 9 m, which implies that the DRIS can transport 1.33 times of jobs in unit time compared to the SRIS. This highest ration can be changed by changing the length of Zones A and C in different lengths.

4 Conclusion

This study has developed an analytical model based on a combined scheduling and simulation approach to measure the performance of the dual robot in-line stocker (DRIS) that is a new type of automated material handling system. A unit DRIS has both processing machines and storing racks inside and uses two robots in the single lane to move more transport jobs in a short time period. The analysis model proposed by this paper can precisely calculate the locations of two robots over time during transport activities. Especially, the model was designed to measure necessary time for waiting or backtracking to avoid collision of two robots. An experimental analysis was conducted to validate the correctness and usefulness of the model based on the data used in an actual manufacturing site.

The analysis model developed by this paper is expected to be used in various ways in practice. First, it can be helpful to determine in which case the DRIS is used and in which case the SRIS would be enough under given input parameters. As illustrated in Sect. 3, if the length of the DRIS is too short or the demand of transport jobs

Table 1 The ratios of disjointed jobs and waiting or backtracking time with the length of the DRIS

Length in meter	5	7	9	11	13	15	17	19	21	23	25	27	29
The ratio of disjointed jobs	1.28	1.16	1.10	1.07	1.05	1.04	1.03	1.02	1.02	1.02	1.01	1.01	1.01
The ratio of waiting or backtracking time	0.80	0.60	0.57	0.62	0.65	0.69	0.70	0.74	0.75	0.76	0.75	0.75	0.74
Capacity ratio of DRIS to SRIS	0.96	1.23	1.33	1.32	1.30	1.28	1.25	1.23	1.20	1.19	1.18	1.16	1.16

is too low, the use of the DRIS might be an excessive investment. Next, the model presented in this paper can be utilized to determine the controllable parameters of a DRIS. As described above, the lengths of Zones A and C were found to have a very large effect on the performance of the DRIS. Depending on a layout inside a DRIS, demand patterns of transport jobs are quite different since a certain type of transport jobs that are characterized by their origins and destinations are more frequent than others. The model can provide quick and precise evaluation results over different layout or zoning alternatives, which is one of the most important advantages of the analytical model [6] as compared to using a commercial simulation software that requires high purchasing and maintenance costs.

References

1. Chung, J., Kim, P.S.: A performance model for stocker systems in liquid crystal display (LCD) fabrication lines. J. Soc. Korea Ind. Syst. Eng. **34**(3), 1–7 (2011)
2. Eldemir, F., Graves, R.J., Malmborg, C.J.: New cycle time and space estimation models for automated storage and retrieval system conceptualization. Int. J. Prod. Res. **42**(22), 4767–4783 (2004)
3. Gu, S.J., Choi, B.K.: Simulation model of TFT-LCD line in-line stocker. In: Proceedings of Conference of Korean Industrial Engineering Society, pp. 288–294 (2008)
4. Ho, Y.C., Su, T.S.: The machine layout within a TFT-LCD bay with a multiple-stacker crane in-line stocker. Int. J. Prod. Res. **50**(18), 5152–5172 (2012)
5. Kim, J.H., Park, J.H., Hong, S.P., Hong, C.H., Park, N.I., Jang, Y.J., Kim, S.I., An, S.J. Moon, I.H.: Optimization of transport zoning in consideration of crane utilization in LCD dual stacker. In: Proceedings of Conference of Korean Industrial Engineering Society. 1991–2013 (2016)
6. Mackulak, G.T., Savory, P.: A simulation-based experiment for comparing AMHS performance in a semiconductor fabrication facility. IEEE Trans. Semicond. Manuf. **14**(3), 273–280 (2001)
7. Roodbergen, K.J., Vis, I.F.: A survey of literature on automated storage and retrieval systems. Eur. J. Oper. Res. **194**(2), 343–362 (2009)
8. Su, T.S., Hwang, M.H.: Efficient machine layout design method with a fuzzy set theory within a bay in a TFT-LCD plant. Proc. Manuf. **11**, 1863–1870 (2017)

A Study on the Optimal Assignment Rule in Parallel Production Systems with Two Special Workers

Xiaowen Zhao, Hisashi Yamamoto, and Jing Sun

Abstract It is possible that workers' error, variation of processing time, lack of parts and machine failure will affect the delivery date of each production process. The consecutive delay in the process can lead to the postponing of manufacturing production. By optimizing the workers' assignments, it is possible to achieve the delivery date of the product and reduce the total expected cost of the product. This study takes a parallel production line as an example and considers the problem of how to optimally assign workers to each process when the total expected cost is minimized by having the same number of processes in each production line.

Keywords Parallel production line · Optimal assignment · Limited-cycle problem

1 Introduction

In the area of production management that supports production activities, addressing production scheduling and line balance issues is essential to improving productivity, reducing production costs, and achieving delivery times. In a serial production line, under uncertain conditions, the results and efficiency of a particular period are often influenced not only by that period but also by the risks that occurred before that period. Considering a standard serial production line as an example, whether a certain process (period) satisfies the delivery date (limit) depends on the state of the previous process (previous period). A problem with this kind of limited-cycle called the limited-cycle multiple production periods problem.

Even the assembly line has been widely used for several decades, we can still benefit from it for reducing processing costs and production time. However, no matter how increased the reliability of the machine on the assembly line is by the

X. Zhao (✉) · H. Yamamoto
Tokyo Metropolitan University, Tokyo, Japan
e-mail: Zhao-xiaowen@ed.tmu.ac.jp

J. Sun
Nagoya Institute of Technology, Nagoya, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_22

developing of technology, human error of the worker, appearing as mistake or failure, is inescapable. It will lead to a delay of whole production processes. Because of the variability of ability of workers, the time of the same processing is different from people to people. In particular, the new recruit is not skillful, so they are considered with a higher probability of delay. As a result, there is no important thing than arranging these new recruits, which is called as untrained-workers in the essay.

The first mathematical formalization of assembly line balancing is established by Salveson [1] 60 years ago, and during the assembly line balancing problem develops, Yamamoto [2] produce the limit-cycle model with multiple periods (*LCMwMP*). This model is set with constraint condition (e.g., the target processing time), which is repeated in every multiple period. If the constraint condition is broken through, an expected cost (e.g., penalty cost) will be increased.

Over the decades, many attempts have been made to address the problem of scheduling in flow shops. The most common objective of scheduling is the minimization of the make span. Many researchers have focused their efforts on the development of heuristics that yield optimal or near-optimal solutions for large-scale problems. Another important objective of scheduling is the minimization of the total flow time (or the sum of job completion times) that results in minimum in-process inventory, and heuristics with this objective have been developed by Liu [3], Rajendran [4].

Depending on whether the constraint condition (target processing time) is reset or not, the *LCMwMP* is divided into reset model and non-reset model. In the field of reset model, a recursive formula for the total expected cost and an algorithm for optimal assignments based on the branch and bound method is proposed by Yamamoto [5].

Zhao [6] proposed an in-line series production line and the optimal assignment problem in the case that the number of people in three groups is 1, 2, and *n*-3, and proposed the locally optimal assignment when an untrained worker is in the first process with constant processing time.

In a global production network environment with many varieties, small quantities, and workers with different production capacities, it is necessary to consider the optimal assignment for different target processing times. Zhao [7] proposed the optimal assignment problem with three groups of workers when target processing time follows the continuous distribution.

In the previous research, we consider an in-line series production line and costs reduced by changing the assignment of untrained workers and well-trained workers. In this research, we consider a parallel production line that is closer to the actual production site and can handle complicated situations. We propose an optimal assignment problem of three types of workers with minimum expected cost when the number of processes in each line is the same.

This paper is written as the following prescribed order. First, the reset model as a simple model of *LCMwMP* is introduced.

Then, the propositions of optimal assignment are demonstrated by derivation. Finally, the assignment optimization rules about minority untrained or well-trained workers are discussed by numerical analyses.

2 Model Explanation

In this chapter, we consider a "Reset model" which is a simple model of the *LCMwMP*.

2.1 Reset Model of LCMwMP

The model is considered based on the following definitions by Zhao [7] in Fig. 1 (Figure 1 shows an image of the relationship between the expected cost and the working time for one series production line constituting a parallel):

(1) In a parallel production line system, we consider a production line with m parallel lines of n processes.
(2) For $j = 1,2,\ldots,m$ and $k = 1,2,\ldots,n$, the processes (j,k) represent the process k of the production line j. For example, the process (3,2) represents the process 2 of the production line 3.
(3) The production (we call it job in the following article) is processed in a rotation of process 1, process 2, …, process n of each line and m productions will be processed by all mn processes
(4) Z is the cycle time of all of the processes, which can be also considered as a target processing time. All of the jobs should be accomplished in current process and moved to next process by time Z.
(5) However, because of the various processing abilities of workers, the actual processing time may not always obey the limit of target processing time Z. Idle and delay should be also concerned in this model. For $1 \le w \le n$, the processing time of process w is donated by T_w.

In this model, a regular processing cost $C_t(>0)$ per unit time will permanently occur during a target processing time Z, regardless whether it is idle or delay. It is

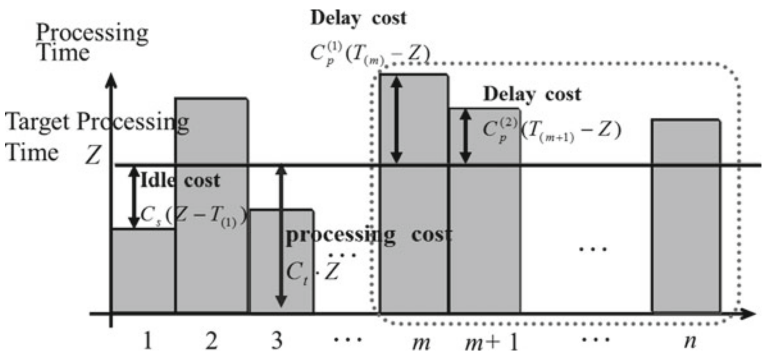


Fig. 1 Description of idle and delay cost in reset model of *LCMwMP*

for the reason that although the job is accomplished prematurely in current process, the next process may be occupied by another job. The job must wait for its start. As a result, an idle cost per unit time, $C_s(\geq 0)$, arises. On the other hand, if the processing time is greater than Z, it is supposed that the delay of process time can be recovered by the overtime works or spare workers in this process, and as a result overtime works or additional resources will be requested in order to meet the target time Z. Thus, a delay cost per unit time, $C_P^{(h)}(\geq 0)$, arises (that is why we call the model a "Reset model").

As a summary of above, we suppose the following costs:

(6) The processing cost per unit time, $C_t(>0)$, for the target processing time limit occurs in each process.
(7) When $T_w < Z$, the idle cost per unit time, $C_s(\geq 0)$ occurs.
(8) When $T_w > Z$, the delay cost per unit time, $C_P^{(h)}(\geq 0)$, occurs in the process if delay occurs in consecutive h processes before its process, for $h = 1, 2, 3, \ldots, n$. If the delay continues for several processes, it can be considered that it will cost more for recover the delay. It is supposed that $C_P^{(h)}$ is increasing in h, which can be expressed as $0 < C_P^{(1)} < C_P^{(2)} < \cdots < C_P^{(n)}$.

2.2 The Assumption of Processing Abilities of Workers

This paper aims to search the optimal assignment of workers for different processing speed, so the reasonable assumption of the property of workers is particularly important. The assumption is the following:

(1) Only one worker can be assigned to each process on each line, for a total of mn workers. Each process must be assigned with one worker.
(2) The processing time of workers is self-dependent. The processing ability is decided by the property of worker own and is not influenced by the processing status such as idle or delay.
(3) In this paper, the workers are distinguished into three types of workers by the processing ability, marking as A, B and C. Worker A (1 workers) represents untrained worker whose processing ability is lower than others, C (1 worker) represents well-trained worker and B (mn -2 workers) represents the general worker. In this paper, we call the untrained worker and the well-trained worker the special workers.
(4) l represents the workers in a production each line. For $l = 1,2,\ldots,m$. These workers have different probability of idle or delay. If the processing time of worker is marked as T_l, where $l \in \{A, B, C\}$.

P_l: The probability of worker l becoming idle, which is $Pr\{T_l \leq Z\}$,
Q_l: The probability of the worker l becoming delayed, which is $Pr\{T_l > Z\}$,
TS_l: The expected idle time of the worker l, which is $E[(Z - T_l)I(T_l \leq Z)]$,
TL_l: The expected delay time of the worker l, which is

$E[(T_l - Z)I(T_l > Z)]$,where $I(O)$ is an index function and given as follows:

$$I(O) = \begin{cases} 1 & (O \text{ is true}) \\ 0 & (O \text{ is not true}). \end{cases}$$

2.3 Optimal Assignment Problem Under Reset Model

One of the most critical problems is how to allocate workers to processes for minimizing the expected cost. We call such a problem with the optimal assignment problem. For describing the optimal assignment problem, we define the following notations:

For $1 \le j \le m$, $1 \le k \le n$,

$A_{(j,k)}$: Worker *A* is assigned in process *k* on line *j*.

$C_{(j,k)}$: Worker *C* is assigned in process *k* on line *j*.

$\pi_{mn}\left(A_{(j,k)}, C_{(j,k)}\right)$: Worker *A* is assigned to $A_{(j,k)}$, worker *C* is assigned to $C_{(j,k)}$, and the other *mn*-2 workers *B* are assigned to the other processes.

$TC\left(mn; \pi_{mn}\left(A_{(j,k)}, C_{(j,k)}\right)\right)$: The total costs of processes 1 to *mn* when workers are allocated by assignment $\pi_{mn}\left(A_{(j,k)}, C_{(j,k)}\right)$ with target processing time *Z*. Which can be expressed as

$$TC\left(mn; \pi_{mn}\left(A_{(j,k)}, C_{(j,k)}\right)\right) = mnC_t Z + f\left(mn; \pi_{mn}\left(A_{(j,k)}, C_{(j,k)}\right)\right) \quad (1)$$

where,

$f\left(mn; \pi_{mn}\left(A_{(j,k)}, C_{(j,k)}\right)\right)$: The sum of the expected idle cost and the expected delay cost caused in process *mn* with target processing time *Z*.

By using these notations, the optimal assignment problem with multiple periods becomes the problem of obtaining an assignment in the following equation:

$$TC\left((mn; \pi^*)\right) = \min_{\pi} TC\left(mn; \pi_{mn}\left(A_{(j,k)}, C_{(j,k)}\right)\right) \quad (2)$$

In this paper, we call π^* the optimal assignment.

3 Preliminary Analysis in a Parallel Production Line

In this paper, we assume that three types of workers, one untrained worker *A*, one well-trained worker *C* and *mn*-2 general workers *B*, depending on their processing ability.

In this chapter, we present the results of a preliminary analysis of the optimal assignment for $m = 2$ and $n = 2$ or 3 when the parallel production line has 2 lines 2 columns and 2 lines 3 columns. We propose the optimal assignment propositions of workers as Proposition 1, Proposition 2 and Proposition 3.

3.1 When the Production Line Has 2 Lines 2 Columns ($m = 2, n = 2$)

In this section, the preliminary analysis results regarding the optimal assignment are described for the case where the production line has 2 lines 2 columns ($m = 2$, $n = 2$).

Proposition 1:

(1) $C_P^{(i)}$ is increasing in i,
(2) $Q_A > Q_B$ and $TL_B > TL_C$,

Then $\pi_{2*2}\left(A_{(1,1)}, C_{(1,2)}\right)$ is the optimal assignment.

As shown in Fig. 2, there are two different cases (a) and (b) where special workers A and C are assigned on the same and different lines. The previous research has shown that the expected cost of a production line is minimized when worker with lower processing speed (untrained worker) is assigned in the first process. Worker with higher processing speed (well-trained worker) is assigned in the last process. Therefore, Fig. 2 shows the optimal assignment when special workers are assigned in the same and different lines.

Proof of Proposition 1:

To prove the Proposition 1, we should to show that

$$f\left(4; \pi_4\left(A_{(1,1)}, C_{(2,2)}\right)\right) - f\left(4; \pi_4\left(A_{(1,1)}, C_{(1,2)}\right)\right) > 0 \tag{3}$$

holds. Therefore,

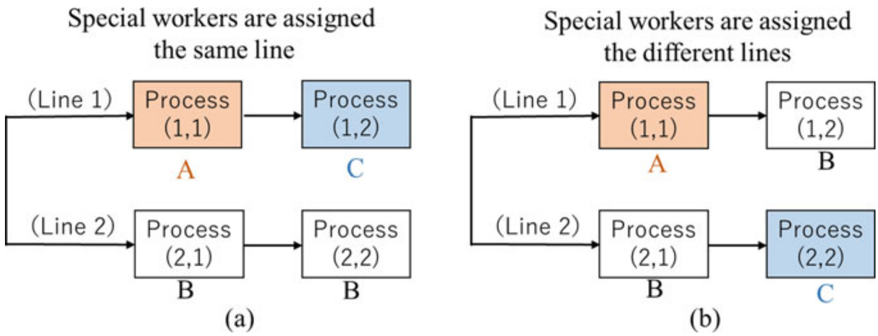


Fig. 2 The optimal assignment of 2 lines 2 columns

$$
\begin{aligned}
& f\left(4; \pi_4\left(A_{(1,1)}, C_{(2,2)}\right)\right) - f\left(4; \pi_4\left(A_{(1,1)}, C_{(1,2)}\right)\right) \\
= & \left(C_P^{(2)} - C_P^{(1)}\right)(Q_A - Q_B)(TL_B - TL_C) \qquad (4)
\end{aligned}
$$

holds. Here, Eq. (4) is positive according to conditions (1) and (2) of Proposition 1.
Proposition 1 is proved.

3.2 When the Production Line Has 2 Lines 3 Columns ($m = 2, n = 3$)

In this section, the preliminary analysis results regarding the optimal assignment are described for the case where the production line has 2 lines 3 columns ($m = 2$, $n = 3$). The proofs of Proposition 2 and Proposition 3 are abbreviated.

Proposition 2:

(1) $C_P^{(i)}$ is increasing in i,

(2) $Q_A > Q_B > Q_c, TL_B/Q_B > TL_C/Q_C$ and $\frac{\frac{1}{Q_B}\cdot\frac{TL_C}{Q_C} - \frac{1}{Q_C}\cdot\frac{TL_B}{Q_B}}{\frac{TL_C}{Q_C} - \frac{TL_B}{Q_B}} - \frac{1}{Q_A} > \frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}}$,

Then $\pi_{2*3}\left(A_{(1,1)}, C_{(1,2)}\right)$ is the optimal assignment.

Proposition 3:

(1) $C_P^{(i)}$ is increasing in i,

(2) $Q_A > Q_B > Q_c,\ TL_B/Q_B > TL_C/Q_C$ and $\frac{\frac{1}{Q_B}\cdot\frac{TL_C}{Q_C} - \frac{1}{Q_C}\cdot\frac{TL_B}{Q_B}}{\frac{TL_C}{Q_C} - \frac{TL_B}{Q_B}} - \frac{1}{Q_A} < \frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}}$

Then $\pi_{2*3}\left(A_{(1,1)}, C_{(1,3)}\right)$ is the optimal assignment.

Figure 3 shows four cases where special workers A and C are assigned on the same line (c), (d) and on different lines (e) and (f).

4 Numeral Experiment

Note that the propositions mentioned above is always holding for any processing time distribution. However, we should assume a processing time distribution when we execute the numeral experiment.

In this chapter, the results of the numeral experiment are described for Proposition 1, Proposition 2 and Proposition 3.

It is assumed that the processing time of these three kinds of workers follows the Exponential distribution. The probability of the processing time probability density function of the worker l is

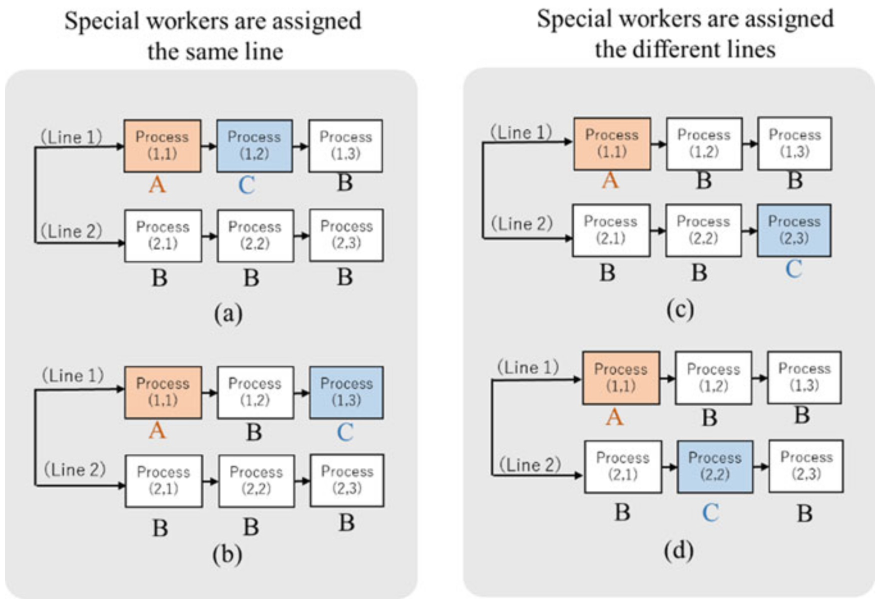


Fig. 3 The optimal assignment of 2 lines 3 columns

$$f_l(t) = \mu_l e^{-\mu_l t} \tag{5}$$

Here, μ_l is the processing speed parameter, as μ_l is bigger, processing speed is faster and lower possibility of delay of processing. From the above, we know that for $l_1, l_2 \in \{A, B, C\}$, if $\mu_{l_1} < \mu_{l_2}$, then $Q_{l_1} > Q_{l_2}$ and $TL_{l_1}/Q_{l_1} > TL_{l_2}/Q_{l_2}$. That means in the case of an exponential distribution, the conditions of Proposition 1, Proposition 2 and Proposition 3 are expressed in terms of the process speeds.

In this chapter, we perform numerical experiments when μ_A, μ_B and μ_C are changed. Tables 1 and 2 show the optimal assignment and the minimum expected cost when the special workers are assigned the same line and different lines. Comparing the expected costs in the above two cases, the lowest total expected cost is shown in red.

Form Table 1, as a result of Proposition 1, it was confirmed that assignment (a) is the optimal assignment.

From Table 2, as a result of Proposition 2 and Proposition 3, they were confirmed that the assignment (c) and the assignment (d) are the optimal assignment. In particular, the optimal assignment is (d) when μ_A, μ_B and μ_C are small. This result coincides with condition (2) of Proposition 3.

Table 1 Optimal assignment and minimum expected cost when the production line has 2 lines 2 columns ($m = 2, n = 2$)

μ_A	μ_B	μ_C	Optimal assignment	Minimum expected cost of the same line	Minimum expected cost of the different line	μ_A	μ_B	μ_C	Optimal assignment	Minimum expected cost of the same line	Minimum expected cost of the different line	μ_A	μ_B	μ_C	Optimal assignment	Minimum expected cost of the same line	Minimum expected cost of the different line
0.1	0.9	1.0	(a)	502.6	503.9	0.1	0.2	1.0	(a)	815.9	834.9	0.1	0.2	0.3	(a)	926.3	953.3
0.2	0.9	1.0	(a)	311.7	312.6	0.1	0.3	1.0	(a)	650.1	668.4	0.1	0.2	0.4	(a)	877.5	890.7
0.3	0.9	1.0	(a)	253.0	253.7	0.1	0.4	1.0	(a)	578.8	539.4	0.1	0.2	0.5	(a)	851.5	867.1
0.4	0.9	1.0	(a)	226.8	227.3	0.1	0.5	1.0	(a)	542.9	553.7	0.1	0.2	0.6	(a)	836.5	853.4
0.5	0.9	1.0	(a)	213.5	213.5	0.1	0.6	1.0	(a)	523.4	531.0	0.1	0.2	0.7	(a)	827.4	845.2
0.6	0.9	1.0	(a)	205.4	205.6	0.1	0.7	1.0	(a)	512.4	517.4	0.1	0.2	0.8	(a)	821.7	840.1
0.7	0.9	1.0	(a)	200.8	201.0	0.1	0.8	1.0	(a)	506.1	509.0	0.1	0.2	0.9	(a)	818.1	836.9
0.8	0.9	1.0	(a)	198.2	198.2	0.1	0.9	1.0	(a)	502.6	503.9	0.1	0.2	1.0	(a)	815.9	834.9

Table 2 Optimal assignment and minimum expected cost when the production line has 2 lines 3 columns ($m = 2, n = 3$)

μ_A	μ_B	μ_C	Optimal assignment	Minimum expected cost of the same line	Minimum expected cost of the different line	μ_A	μ_B	μ_C	Optimal assignment	Minimum expected cost of the same line	Minimum expected cost of the different line	μ_A	μ_B	μ_C	Optimal assignment	Minimum expected cost of the same line	Minimum expected cost of the different line
0.1	0.9	1.0	(c)	604.43	606.0	0.1	0.2	1.0	(c)	1396.2	1436.6	0.1	0.2	0.3	(d)	1638.1	1650.2
0.2	0.9	1.0	(c)	413.19	414.4	0.1	0.3	1.0	(c)	966.7	1001.3	0.1	0.2	0.4	(d)	1562.4	1580.2
0.3	0.9	1.0	(c)	354.29	355.2	0.1	0.4	1.0	(c)	788.0	813.0	0.1	0.2	0.5	(c)	1514.2	1541.7
0.4	0.9	1.0	(c)	327.86	328.5	0.1	0.5	1.0	(c)	700.5	717.5	0.1	0.2	0.6	(c)	1475.6	1507.2
0.5	0.9	1.0	(c)	314.0	314.0	0.1	0.6	1.0	(c)	653.8	664.9	0.1	0.2	0.7	(c)	1447.1	1481.8
0.6	0.9	1.0	(c)	306.15	306.5	0.1	0.7	1.0	(c)	627.8	634.5	0.1	0.2	0.8	(c)	1425.6	1462.7
0.7	0.9	1.0	(c)	301.52	301.7	0.1	0.8	1.0	(c)	612.9	616.6	0.1	0.2	0.9	(c)	1409.1	1448.0
0.8	0.9	1.0	(c)	298.77	298.9	0.1	0.9	1.0	(c)	604.4	606.0	0.1	0.2	1.0	(c)	1369.2	1436.6

5 Conclusion

In this paper, we focused on the parallel production line with 2 lines 2 columns and 2 lines 3 columns. We proposed an optimal assignment that minimizes the total expected costs using a limited-cycle model with multiple periods when special workers are assigned the same line. We also considered through the numerical experiment of the optimal assignment law other than the conditions of the proposed optimal assignment rule.

As a future research, we propose an optimal assignment rule of parallel production line with m lines and n columns. In addition, we also consider an optimal assignment when the different number of processes on each production line.

References

1. Salveson, M.E.: The assembly line balancing problem. J. Ind. Eng. **6**(3), 18–25 (1955)
2. Yamamoto, H., Matsui, M., Liu, J.: A basic study on a limited-cycle problem with multi periods and the optimal assignment problem. J. Jpn. Ind. Manage. Assoc. **57**(1), 23–31 (2006)
3. Liu, J., Reeves, C.R.: Constructive and composite heuristic solutions to the P//ΣCi scheduling problem. Eur. J. Oper. Res. **132**(2), 439–452 (2001)
4. Rajendran, C., Ziegler, H.: Ant colony algorithms for permutation flowshop scheduling to minimize makespan/total flowtime of jobs. Eur. J. Oper. Res. **155**(2), 426–438 (2004)
5. Yamamoto, H., Matsui, M., Bai, X.S.: A branch and bound method for the optimal assignment during a Limit-cycle problem with multiple periods. J. Jpn Ind. Manage. Assoc. **58**(1), 37–43 (2007)
6. Zhao, X., Yamamoto, H., Sun, J.: Optimal assignment with three special workers in a reset limited-cycle problem with multiple periods. In: The Fourteenth International Conference on Industrial Management, Hangzhou, China. (CD-ROM) 2018
7. Zhao, X., Yamamoto, H. and Sun, J.: Optimal assignment of three special workers in reset limited-cycle multiple periods with variable target processing time. In: The 19th Asia Pacific Industrial Engineering and Management System Conference, Hong Kong (2018)

Considering a Deteriorating EOQ Model Under Stochastic Demand and Shortage Allowed

Chiang-Sheng Lee, Chien-Hung Cho, Cheng-Thai Tsai, and Hsine-Jen Tsai

Abstract The inventory model for deteriorating items with a stochastic demand rate is studied in our paper. Inventory models dealing with the deterioration of items have attracted considerable interests in recent decades since the deteriorating phenomenon is a crucial factor affecting the profits of a company. However, most of current inventory models considering deterioration assumes a certain or a constant demand function which is unreasonable in the prevailing market that conforms to our daily reality. Therefore, stochastic demand must be considered. Besides, shortage is allowed in the study. We provide approximating solutions for optimal ordering quantities for our model, and show that our optimal solutions from the assumed models give conditions and results very close to the optimal solutions obtained by computation. Further, these results illustrate the impact of various parameters on the optimal policy and the profit.

Keywords EOQ model · Deteriorating EOQ model · Stochastic demand

1 Introduction

Deterioration is a very common phenomenon in our daily life, such as: metamorphism for medicine and film, volatility for packaged chemicals, decomposition for food, and functional degradation for whole-blood unit, etc. In recent decades, inventory models dealing with the deterioration items have attracted considerable interests since the deteriorating phenomenon is a crucial factor affecting the profits of a company.

C.-S. Lee (✉) · C.-H. Cho · C.-T. Tsai
National Taiwan University of Science and Technology, Taipei, Taiwan
e-mail: cslee@mail.ntust.edu.tw

H.-J. Tsai
Fu-Jen Catholic University, New Taipei City, Taiwan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_23

Because of the page limitation, we will omit the literature review. For the detailed literatures, we refer readers to [1, 2]; [1] gave an advanced review of the deteriorating inventory literatures, which including 130 references, from the early 1990s to 1999s, and [2] reviewed literatures for deteriorating inventory from 2001 to 2011 including 240 references.

In the study, we investigate the strategy of optimal ordering quantity of the inventory system for a single deteriorating item. Under the condition of stochastic demand with fixed shelf lifetime, we consider the inventory policy that shortage is allowed and the items are not backlogged even after the stock depletes. Only at the expiration date can the replenishment arrived instantaneously.

2 Assumptions, Notations and Model Descriptions

2.1 Assumptions

The following assumptions are used in this study:

1. Time horizon is infinite.
2. The replenishment rate is instantaneous and the lead time is zero.
3. Replenishment and repair are not allowed in deteriorating items.
4. The demand in each cycle is following uniform distribution from A to B.
5. Shortage is allowed in the model.
6. Considers single items with a fixed price.
7. Determination of parameters and expiration date are established.

2.2 Notations

The definitions of all notations used in this paper are described as following:

A Minimum demand rate in the period of T;
B Maximum demand rate in the period of T;
δ The demand that satisfies the inventory reaches the zero level at time t when $\delta \in [A, B]$;
X Stochastic demand rate follows uniform distribution with parameter A to B in the period of T;
θ The coefficient of deterioration;
$Q(\delta)$ The order quantity with a constant demand rate δ;
P The selling price per unit;
K The fixed order cost;
H The holding cost per unit per unit time;

S	The shortage cost per unit;
K_1	The deterioration cost per unit;
K_2	The expired cost per unit;
T	The expiration date of a single item;
t_1	The time at which the inventory drops to zero;
$I_X(t)$	The inventory level at time t with stochastic demand rate X at time t, $0 \leq t \leq T$.

2.3 Model Descriptions

Model: Shortage is allowed before the expiration date T, and the shortage would occur within the time from t_1 to T. Only at T can the replenishment be arrived instantaneously.

Under the stochastic demand, we investigate a demand rate δ for maximum total relevant profit such that the inventory will reach to zero at that $I_\delta(T) = 0$. Therefore, two cases are described for the model at below.

Case I (X > δ): Because of high demand, the inventory drops to zero before the expiration date T. In a simple word, the amount of demand is greater than δ demand in the period of T. Overall effect of deterioration and high demand would be certainly greater than the ordering quantity.

Case II (X ≤ δ): During the period of weak demand, the remainder of the inventory exists at the expiration date T. On the other hand, the amount of demand is smaller than δ demand in the period of T. Overall effect of deterioration and weak demand would be certainly smaller than the ordering quantity. Particularly, in this case, the remainder must be discarded with cost.

3 Mathematical Model

During the interval [0, T], the deterioration inventory depletion is influenced by the random demand X and deterioration rate $\theta I_X(t)$.The model considered in our study that represents the inventory level at time t can be descripted by:

$$\frac{\partial I_X(t)}{\partial t} = -X - \theta I_X(t) \quad 0 \leq t \leq T \tag{1}$$

where θ is the coefficient of deterioration and $X \sim uniform(A, B)$. After computation, the general solution to the differential equation (1) is $I_X(t) = \frac{cX}{\theta}e^{-\theta t} - \frac{X}{\theta}$ for some value c.

First, the objective of the model is to determine the ordering quantity at time zero for items having stock with independent demand rate δ, and with a constant coefficient of decay θ. By using first order linear differential equation, we obtain an equation of inventory level $I_\delta(t) = \frac{\delta}{\theta}\left[e^{\theta(T-t)} - 1\right]$ which is based on two conditions: (1) the ordering quantity is $Q(\delta) = I_\delta(0)$ when the demand rate X equals δ at the beginning time and (2) the inventory depletes at T, or $I_\delta(T) = 0$. Thus the ordering quantity $Q(\delta)$ is obtained as following.

$$Q(\delta) = I_\delta(0) = \frac{\delta}{\theta}\left[e^{\theta T} - 1\right], \quad A \leq \delta \leq B \tag{2}$$

For any single item, the ordering quantity $Q(\delta)$ would be influenced by expiration date T and the coefficient of deterioration θ. The solution of Eq. (1) is solved by first order linear differential equation, under the conditions: (1) at the boundary condition of time zero $I_X(0) = Q(\delta)$, and (2) at the boundary condition of time T, $I_\delta(T) = 0$, so that we obtain the inventory level

$$I_X(t) = \left[Q(\delta) + \frac{X}{\theta}\right]e^{-\theta t} - \frac{X}{\theta} \tag{3}$$

Therefore, the inventory cycle of Case I and Case II are demonstrated distinctively for the model as following.

3.1 *The Model*

In this model, the inventory level depleting before time T occurs shortage during the period of outstanding demand, therefore, it is assumed that the demand during the time from $t_1 < t < T$ has a random demand X. Before formulating mathematics, the model can be discussed in the following 2 cases.

Case I: $X > \delta$

In Case I, the point t_1 in which the inventory level reaches to zero is earlier than time T, since at the beginning of the cycle, $Q(\delta)$ is not sufficient for the period of demand rate between δ and B. The point t_1 is represented in Eq. (4) and boundary of t_1 is shown in Eq. (5).

$$t_1 = \frac{1}{\theta}\ln\left(1 + \frac{\delta\left[e^{\theta T} - 1\right]}{X}\right) = \frac{1}{\theta}\ln\left(1 + \frac{\theta Q(\delta)}{X}\right) \tag{4}$$

The boundary region of t_1 as $\min(t_1) \leq t_1 \leq$ T

$$\Rightarrow \frac{1}{\theta}\ln\left(\frac{\theta Q(\delta)}{B}+1\right) \le t_1 \le \frac{1}{\theta}\ln\left(\frac{\theta Q(\delta)}{\delta}+1\right) = T, \qquad if\ \delta \le X \le B \quad (5)$$

And the shortage $X(T - t_1)$ units will be calculated in Eq. (11).

Case II: $X \le \delta$

In Case II, the unsalable items is discarded with $G(\delta)$ units, and the inventory level goes up to $Q(\delta)$ at time T instantaneously. Thus, the equation for G(δ) can be written as

$$G(\delta) = I_X(T) = \left[\frac{\delta - X}{\theta}\right]\left[1 - e^{-\theta T}\right] = \left[\frac{Q(\delta)}{e^{T\theta} - 1} - \frac{X}{\theta}\right]\left[1 - e^{-\theta T}\right], \quad if\ X < \delta \quad (6)$$

The followings are the deterministic cost and profit for the total relevant profit within two cases are established. The ordering cost per cycle $= K$

The purchasing cost per cycle $= CQ(\delta)$

$$\text{The selling price per cycle} = \begin{cases} PXt_1,\ if\ X > \delta \\ PXT,\ if\ X \le \delta \end{cases} \quad (7)$$

$$\text{The holding cost per cycle} = \begin{cases} H\left[\frac{Q(\delta)}{\theta} - \frac{Xt_1}{\theta}\right], & if\ X > \delta \\ H\left[\frac{1}{\theta}\left(Q(\delta) + \frac{X}{\theta}\right)\left(1 - e^{-\theta T}\right) - \frac{XT}{\theta}\right], & if\ X \le \delta \end{cases} \quad (8)$$

$$\text{The deterioration cost per cycle} = \begin{cases} K_1[Q(\delta) - Xt_1], & if\ X > \delta \\ K_1\left[Q(\delta) - Q(\delta - X)e^{-\theta T} - XT\right], & if\ X \le \delta \end{cases} \quad (9)$$

$$\text{The outdating cost per cycle} = K_2\left[\left[\frac{Q(\delta)}{e^{T\theta} - 1} - \frac{X}{\theta}\right]\left[1 - e^{-\theta T}\right]\right], \quad if\ X \le \delta \quad (10)$$

$$\text{The shortage cost in the model is} = SX(T - t_1) \quad (11)$$

Hence, the total relevant expected profit can be written as:

$$ETRP(Q(\delta)) = \int_{\delta}^{B} \frac{1}{T}\Bigg(Pxt_1 - \Bigg(K + \frac{H}{\theta}(Q(\delta) - xt_1) + K_1 * (Q(\delta) - xt_1)$$
$$+ CQ(\delta) + Sx(T - t_1))) * f(x)dx$$
$$+ \int_{A}^{\delta} \frac{1}{T}\Bigg(PxT - \Bigg(K + \frac{H}{\theta}\left(\left[Q(\delta) + \frac{x}{\theta}\right] * \left[1 - e^{-\theta T}\right] - xT\right)$$

$$
\begin{aligned}
&+ K_1 * (Q(\delta) - xT) + [K_2 - K_1] \\
&* \left[\frac{\delta}{\theta} - \frac{x}{\theta}\right] * \left[1 - e^{-\theta T}\right] + CQ(\delta)\Big)\Big) \\
&* f(x)dx = \frac{1}{[B - A]}\left\{\left(P + \frac{H}{\theta} + K_1 + S\right)\right. \\
&* \left(\frac{B\theta Q(\delta) + B^2 \ln\left(1 + \frac{\theta Q(\delta)}{B}\right)}{2T\theta}\right. \\
&\left. - \frac{\frac{\theta^2 Q(\delta)^2}{\left[e^{T\theta} - 1\right]} + \theta^2 Q^2\left(\ln 0\left(\frac{B}{Q(\delta)\theta} + 1\right) + \ln\left(\left[e^{T\theta} - 1\right]\right) - T\theta\right)}{2T\theta}\right) \\
&- \frac{\left(\frac{H}{\theta}\right)\left(BQ - \frac{Q(\delta)^2\theta}{\left[e^{T\theta} - 1\right]} + \left[\left(\frac{Q(\delta)^2\theta}{\left[e^{T\theta} - 1\right]} - AQ\right) + \frac{\left[\frac{Q(\delta)^2\theta^2}{\left[e^{T\theta} - 1\right]^2} - A^2\right]}{2\theta}\right]\left[1 - e^{-\theta T}\right]\right)}{T} \\
&- \frac{SB^2}{2} - \frac{\left(P + \frac{H}{\theta} + K_1\right)A^2}{2} - \frac{(K + K_1 Q(\delta) + CQ(\delta))(B - A)}{T} \\
&\left. - \frac{[K_2 - K_1]\frac{\left[\frac{Q(\delta)^2\theta^2}{\left[e^{T\theta} - 1\right]^2} - \frac{2Q\theta A}{\left[e^{T\theta} - 1\right]} + A^2\right]}{2\theta}\left[1 - e^{-\theta T}\right]}{T}\right\}
\end{aligned} \tag{12}
$$

3.2 Solution Process for the Model

Next, our goal is to obtain the partial derivatives $\frac{dETRP}{dQ(\delta)}$.

$$
\begin{aligned}
\frac{dETRP}{dQ(\delta)} = \frac{1}{[B - A]}&\left\{\left(P + \frac{H}{\theta} + K_1 + S\right)\right. \\
&* \left(\frac{B}{T} - \frac{Q(\delta)\theta \ln\left(\frac{Q(\delta)\theta}{B} + 1\right)}{T}\right. \\
&+ \frac{Q(\delta)\theta \ln\left(\frac{Q(\delta)}{B}\right)}{T} + \frac{Q(\delta)\theta \ln(\theta)}{T}
\end{aligned}
$$

$$- \frac{Q(\delta)\theta \ln\left(e^{T\theta} - 1\right)}{T} + Q(\delta)\theta^2 - \frac{Q(\delta)\theta}{T\left[e^{T\theta} - 1\right]}\Bigg)$$

$$- \frac{\left(\frac{H}{\theta}\right)\left(B - \frac{2Q(\delta)\theta}{\left[e^{T\theta}-1\right]}\right)}{T} - \frac{(K_1 + C)(B - A)}{T}$$

$$- \frac{\frac{H}{\theta}\left[\left(\frac{2Q(\delta)\theta}{\left[e^{T\theta}-1\right]}\right) + \frac{Q(\delta)\theta}{\left[e^{T\theta}-1\right]^2} - A\right]\left[1 - e^{-\theta T}\right]}{T}$$

$$- \frac{\left[K_2 - K_1\right]\left[\frac{Q(\delta)\theta}{\left[e^{T\theta}-1\right]} - A\right]e^{-\theta T}}{T}\Bigg\} \tag{13}$$

Because of the character of natural logarithm, in Eq. (13), it is difficult to obtain closed forms for $\ln\left(\frac{Q(\delta)\theta}{B} + 1\right)$ and $\ln\left(\frac{Q(\delta)}{B}\right)$, we utilize Taylor series for the approximations of $\ln\left(\frac{Q(\delta)\theta}{B} + 1\right)$ and $\ln\left(\frac{Q(\delta)}{B}\right)$. Let

$$\ln(u) = \ln\left(\frac{1+y}{1-y}\right) = 2y\left(1 + \frac{1}{3}y^2 + \frac{1}{5}y^4 + \frac{1}{7}y^6 \cdots\right),$$

where $y = \frac{u-1}{u+1}$. We extract the first one in the terms of Taylor series to be our value of approximation as shown in below.

$$\ln\left(\frac{Q(\delta)\theta}{B} + 1\right) \approx \frac{2Q(\delta)\theta}{Q(\delta)\theta + 2B} \tag{14}$$

$$\ln\left(\frac{Q(\delta)}{B}\right) \approx \frac{2(Q(\delta) - B)}{(Q(\delta) + B)} \tag{15}$$

Hence, the approximation of Eq. (13) can be obtained as:

$$\begin{aligned}
\frac{dETRP}{dQ} = &\left(P + \frac{H}{\theta} + K_1 + S\right) * \left\{BQ(\delta)^2\theta + Q(\delta)^3\theta^2 \ln(\theta)\right.\\
&- Q(\delta)^3\theta^2 \ln\left(e^{T\theta} - 1\right) + TQ(\delta)^3\theta^3 - \frac{Q(\delta)^3\theta^2}{e^{T\theta} - 1}\\
&+ B^2Q(\delta)\theta - 4Q(\delta)^2\theta^2 B + BQ(\delta)^2\theta^2 \ln(\theta)\\
&- BQ(\delta)^2\theta^2 \ln\left(e^{T\theta} - 1\right) + BTQ(\delta)^2\theta^3\\
&- \frac{BQ(\delta)^2\theta^2}{\left[e^{T\theta} - 1\right]} + 2B^2Q(\delta) + 2BQ(\delta)^2\theta \ln(\theta)\\
&- 2BQ(\delta)^2\theta \ln\left(e^{T\theta} - 1\right) + 2BTQ(\delta)^2\theta^2 + 4Q(\delta)^2\theta B
\end{aligned}$$

$$
\begin{aligned}
&-\frac{2BQ(\delta)^2\theta}{\left[e^{T\theta}-1\right]}+2B^3-4Q(\delta)\theta B^2\\
&+2B^2Q(\delta)\theta\ln(\theta)-2B^2Q(\delta)\theta\ln\left(e^{T\theta}-1\right)\\
&+2B^2TQQ(\delta)\theta^2-\left.\frac{2B^2Q(\delta)\theta}{\left[e^{T\theta}-1\right]}\right\}-\left(\frac{H}{\theta}\right)\\
&\left(BQ(\delta)^2\theta-\frac{2Q(\delta)^3\theta^2+2BQ(\delta)^2\theta^2+4BQ(\delta)^2\theta+4B^2Q(\delta)\theta}{e^{T\theta}-1}\right.\\
&\left.+B^2Q(\delta)\theta+2B^2Q(\delta)+2B^3\right)\\
&-(K_1+C)\left(BQ(\delta)^2\theta-AQ(\delta)^2\theta+B^2Q(\delta)\theta\right.\\
&\left.-Q(\delta)\theta AB+2B^2Q(\delta)-2BAQ(\delta)+2B^3-2AB^2\right)\\
&-\frac{H}{\theta}\left\{\left(\frac{2Q(\delta)^3\theta^2+2BQ(\delta)^2\theta^2+4BQ(\delta)^2\theta+4B^2Q(\delta)\theta}{e^{T\theta}-1}\right)\right.\\
&+\frac{Q(\delta)^3\theta^2+BQ(\delta)^2\theta^2+2BQ(\delta)^2\theta+2B^2Q(\delta)\theta}{\left[e^{T\theta}-1\right]^2}\\
&\left.-AQ(\delta)^2\theta-ABQ(\delta)\theta-2BAQ(\delta)-2AB^2\right\}\left[1-e^{-\theta T}\right]\\
&-[K_2-K_1]\left[\frac{Q(\delta)^3\theta^2+BQ(\delta)^2\theta^2+2BQ(\delta)^2\theta+2B^2Q(\delta)\theta}{e^{T\theta}-1}\right.\\
&\left.-Q(\delta)^2\theta A-ABQ(\delta)\theta-2BQ(\delta)A-2AB^2\right]e^{-\theta T}=0
\end{aligned}
\tag{16}
$$

After the arrangement, Eq. (16) can be expressed by the following Eq. (17).

$$
\begin{aligned}
&Q(\delta)^3\left(\left(P+\frac{H}{\theta}+K_1+S\right)\right.\\
&\left(\theta^2\ln\left(\frac{\theta}{e^{T\theta}-1}\right)+T\theta^3-\frac{\theta^2}{e^{T\theta}-1}\right)+\left(\frac{H}{\theta}\right)\\
&\left(\frac{2\theta^2}{e^{T\theta}-1}-\left[\left(\frac{2\theta^2}{e^{T\theta}-1}\right)+\frac{\theta^2}{\left[e^{T\theta}-1\right]^2}\right]\left[1-e^{-\theta T}\right]\right)\\
&\left.-[K_2-K_1]\left[\frac{\theta^2}{e^{T\theta}-1}\right]e^{-\theta T}\right)\\
&+Q(\delta)^2\left(\left(P+\frac{H}{\theta}+K_1+S\right)\left(B\theta+2B\theta\ln\left(\frac{\theta}{e^{T\theta}-1}\right)\right.\right.\\
&+2BT\theta^2+4\theta B-\frac{2B\theta}{e^{T\theta}-1}-4\theta^2B+B\theta^2\ln\left(\frac{\theta}{e^{T\theta}-1}\right)\\
&\left.+BT\theta^3-\frac{B\theta^2}{e^{T\theta}-1}\right)+\left(\frac{H}{\theta}\right)\left(\frac{4B\theta+2B\theta^2}{e^{T\theta}-1}-B\theta\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{H}{\theta}\left[\frac{4B\theta + 2B\theta^2}{e^{T\theta}-1} + \frac{2B\theta + B\theta^2}{\left[e^{T\theta}-1\right]^2} - A\theta\right]\left[1 - e^{-\theta T}\right] \\
&-[K_2 - K_1]\left[\frac{2B\theta + B\theta^2}{e^{T\theta}-1} - \theta A\right]e^{-\theta T} - (K_1 + C)(B\theta - A\theta)\Bigg) \\
&+ Q(\delta)\left(\left(P + \frac{H}{\theta} + K_1 + S\right)\left(2B^2 - 3\theta B^2 + 2B^2\theta \ln\left(\frac{\theta}{e^{T\theta}-1}\right)\right.\right. \\
&\left.+2B^2T\theta^2 - \frac{2B^2\theta}{e^{T\theta}-1}\right) + \left(\frac{H}{\theta}\right)\left(\frac{4B^2\theta}{e^{T\theta}-1} - 2B^2 - B^2\theta\right) \\
&- \frac{H}{\theta}\left[\left(\frac{4B^2\theta}{e^{T\theta}-1}\right) + \frac{2B^2\theta}{\left[e^{T\theta}-1\right]^2} - 2BA - AB\theta\right]\left[1 - e^{-\theta T}\right] \\
&- [K_2 - K_1]\left[\frac{2B^2\theta}{e^{T\theta}-1} - 2BA - AB\theta\right]e^{-\theta T} \\
&-(K_1 + C)\big(2B^2 - 2BA + B^2\theta - \theta AB\big)\Big) \\
&- \left[\frac{H}{\theta}2B^3 + (K_1 + C)\big(2B^3 - 2AB^2\big) - [K_2 - K_1]2AB^2e^{-\theta T}\right. \\
&\left.-2AB^2\frac{H}{\theta}\left[1 - e^{-\theta T}\right] - \left(P + \frac{H}{\theta} + K_1 + S\right)2B^3\right] = 0 \qquad (17)
\end{aligned}
$$

Then the form of the cubic equation $aQ(\delta)^3 + bQ(\delta)^2 + cQ(\delta) + d = 0$ is expressed, where a, b, c, and d are constant. With the aim of optimal ordering quantity, we derive the solution of approximation $Q^*(\delta^*)$ of the cubic equation in Lemma 1 as follows.

Lemma 1 *If the discriminant of the cubic equation* $\Delta = 18abcd - 4b^3d + b^2c^3 - 4ac^3 - 27ad$ *is negative, then* $Q^*(\delta^*)$ *has one real root and two complex conjugate roots. If* Δ *is positive, then* $Q^*(\delta^*)$ *has three distinct real roots. Moreover, if* Δ *equals to zero, then* $Q^*(\delta^*)$ *has a multiple real roots and one distinct real root. Thus, the solutions can be found from the following computations:*

$$
\Delta_1 = b^2 - 3ac,\ \Delta_2 = 2b^3 - 9abc + 27a^2d,\ \text{and}\ \eta = \sqrt[3]{\frac{\Delta_2 \pm \sqrt{\Delta_2^2 - 4\Delta_1^3}}{2}}
$$

Hence, the general solution of the cubic equation for the ordering quantity of approximation is represented in $Q^*(\delta^*) = -\frac{1}{3a}\left(b + \xi^k\eta + \frac{\Delta_1}{\xi^k\eta}\right),\ k\epsilon\{0, 1, 2\}$ where $\xi = -\frac{1}{2} - \frac{1}{2}\sqrt{3}i$ is a cube root of unity. Therefore, the results of the general solution of the cubic equation solved by Tartaglia are shown in following.

The discriminant of the cubic equation is determined as follows.

$$\Delta = 18abcd - 4b^3d + b^2c^3 - 4ac^3 - 27ad$$

Note that: if $\Delta < 0$, then $Q^*(\delta^*)$ has only one real root.

if $\Delta > 0$, then $Q^*(\delta^*)$ has three distinct real roots.

if $\Delta = 0$, then $Q^*(\delta^*)$ has a multiple roots and one distinct real root.

Hence, the inventory system of allowed shortage with random demand indicates that the solution of approximation has three real roots, two real or one real root.

4 Numerical Examples and Analysis for Percentage Error

In this section we present the approximated solution $Q^*(\delta)$ and the optimal solution $Q^{**}(\delta)$ to investigate the percentage error of the approximated solution ε by setting $\varepsilon = \frac{Q^*(\delta) - Q^{**}(\delta)}{Q^{**}(\delta)} \times 100\%$ to verify the analytical results from the model.

4.1 Analysis for Percentage Error

It follows the base setting of parameters:

$P = 130$, $T = 1$, $\theta = 0.08$, $C = 50$, $K_1 = 2$, $K_2 = 5$, $H = 2$, $K = 300$, $S = 75$, and $(A, B) \in \{U_1, U_2, U_3\}$.

In Fig. 1, it clearly shows that the error of approximated solution is not significantly affected by the fixed order cost and selling price. But there are lager errors occurred under the expiration date and deterioration coefficient. In the model, the process solution is approximated after the first partial derivative.

5 Conclusions

In this paper, we developed a specific deteriorating EOQ Model with stochastic demand and fixed shelf-lifetime. These results show that the approximated solution closes to the optimal ordering quantity by computations, and demonstrated the impacts of various parameters on the optimal policy and the profit. Therefore, based on the basic setting parameters, some conclusions are presented as follows:

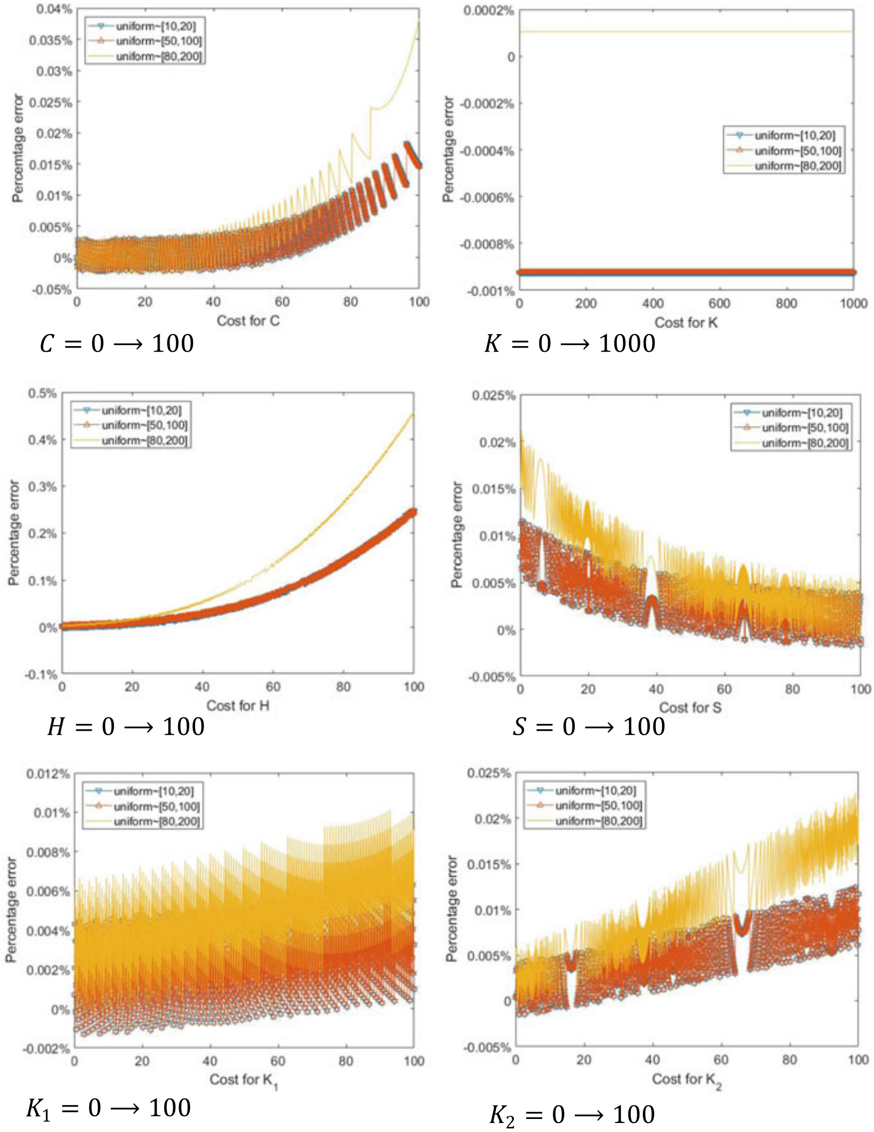


Fig. 1 The percentage error with respect to the parameters in the model

(1) This model has a large percentage error of approximated solution under the increase of the coefficient of deterioration and the expiration date.
(2) In the model with shortage allowed, the parameter δ^* of optimal solution do not significantly close to minimum value A of uniform distribution with the higher index and the expiration date.

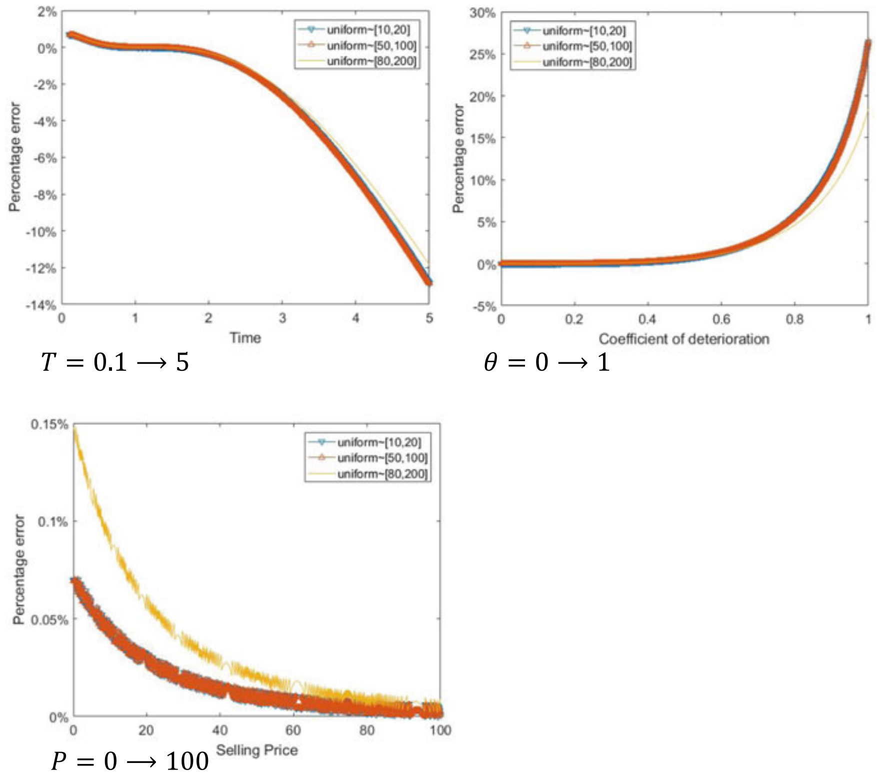


Fig. 1 (continued)

(3) In most of scenarios, the model with shortage allowed is recommended due to the results that the ETRP without shortage is greater than the ETRP with shortage. This model is recommended because the company lose profit very fast under the model without shortage which indicates inventory is depleted rapidly so that replenishment is arrived early.

References

1. Goyal, S.K., Giri, B.C.: Recent trends in modeling of deteriorating inventory. Eur. J. Oper. Res. **16**, 1–16 (2001)
2. Bakker, M., Riezebos, J., Teunter, R.H.: Review of inventory systems with deterioration since 2001. Eur. J. Oper. Res. **221**(2), 275–284 (2012)

Internet of Things, Data Analytics and Smart Manufacturing

Assessing AI-Readiness in Production—A Conceptual Approach

Heidi Heimberger, Djerdj Horvat, and Frank Schultmann

Abstract Due to its high potential to perform many tasks faster, more accurately and in greater detail than humans artificial intelligence (AI) has been attracting growing attention across industries. In manufacturing, AI, in combination with digital sensors, networks and software-based automation, defines a new industrialization age. The integration of AI into production processes promises to boost the productivity, efficiency, as well as the automation of processes. However, AI adoption in manufacturing is currently still in its early stage and lacks practical experiences. This raises the question, to which extent manufacturing companies are ready to implement AI. While approaches to assessing the maturity in terms of the digitalization or Industry 4.0 (I4.0) of manufacturing companies are well established and discussed in the literature, approaches that specifically address AI in manufacturing are still lacking. To address this gap, we present an approach to analyze and monitor the readiness of manufacturing firms for working with AI technologies. In accordance with the existing assessment concepts of digitalization and I4.0, our approach examines different areas of digital technologies on the product and production level of manufacturing firms. Moreover, it incorporates the key foundation for AI—security and data—into a conceptual model. We generally assume that companies need to achieve a certain level of digital readiness in three key dimensions in order to be ready for implementing AI-based technologies. We operationalize these dimensions through a variety of product- and production-specific as well as data- and safety-related indicators. In order to illustrate the implementation of our concept in practical terms, we present the results of the readiness assessment of two German manufacturing companies.

Keywords Artificial intelligence · Readiness · Manufacturing

H. Heimberger (✉) · D. Horvat
Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, Germany
e-mail: heidi.heimberger@isi.fraunhofer.de

H. Heimberger · F. Schultmann
Institute for Industrial Production (IIP), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_24

1 Introduction

The increasing use of digital technologies permeates many areas of life, creating changes in companies [1] and transforming industries [2] by moving from analog to digital processes. In the context of increasing digitization, artificial intelligence (AI) has undergone a resurgence over the last few years [3], attracting attention across industries. In manufacturing, AI combined with digital sensors, networks, and software-based automation is defining a new era of industrialization. Due to its great potential to perform many tasks faster, more accurately, and in greater detail than humans [4], the integration of AI into production processes promises significant performance improvements, e.g. in terms of productivity, flexibility and efficiency [5].

However, the use of AI in manufacturing is currently still in its early stages [3] and lacks practical experience. This raises the question of the extent to which manufacturing companies are ready to implement AI. While approaches to assess the maturity level in terms of Industry 4.0 or the digitalization of manufacturing companies are already established and actively discussed in the literature (e.g., [6–9]), approaches specifically addressing AI in manufacturing are, to the best of our knowledge, scarce. Although a few approaches investigating AI-readiness or maturity have been introduced in the literature [10–13], they do not consider the specifics of manufacturing companies. To address this gap, we present a comprehensive conceptual approach to analyze and assess the readiness of manufacturing companies to use AI technologies. We thereby focus on production and assume that digitization technologies create a technical foundation for the adoption of AI. In addition to the technological domain, many aspects from other areas influence the adoption and use of digital technologies as well, particularly the organization [6–9, 11–14] (e.g., management, strategy, or employees) or the environment [6, 7, 11, 12] (e.g., customers, collaborations, or regulations). In this paper, however, we neglect the non-technological areas guided by the notion that the technological prerequisites representing the basic factors for production should be analyzed first to get a first impression of the AI-readiness of manufacturing companies. Throughout this paper, we use the terms production and manufacturing interchangeably, however, in content we refer to the specificity of manufacturing companies in production.

The paper is structured as follows: After a brief introduction to the conceptual background of our model, we present the dimensions as well as the logic of assessing AI-readiness based on our metric index. To illustrate the implementation of our concept in practical terms, we then present the results of the readiness assessment of two German manufacturing companies. In the closing section, we highlight conclusions as well as the main implications for further research.

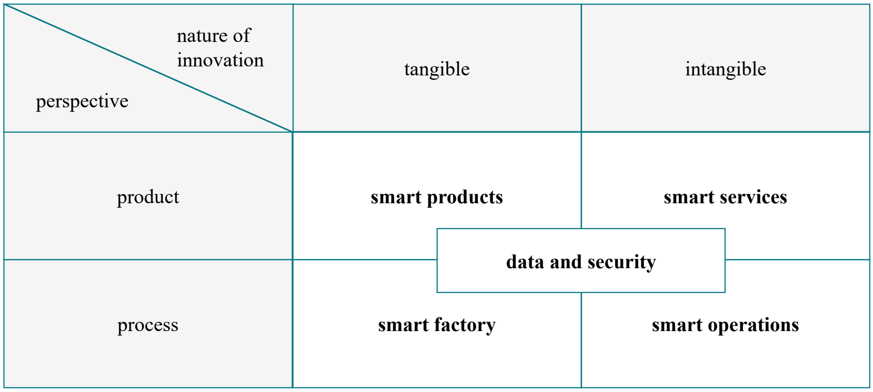


Fig. 1 Digital innovation fields in manufacturing based on [15–17]

2 AI-Readiness Model

2.1 Conceptual Background

The basic conceptual background for our model is the innovation typology for manufacturing companies [15–17]. In general, innovation activities are clustered into two perspectives representing product and process innovations, which in turn can take tangible or intangible form. Considering these dimensions in the sense of digital value creation, a matrix of four domains emerges: smart factory, smart products, smart operations, and smart services [16, 17]. In addition to these four domains within the product and process perspective, scholars and practitioners agree that the availability of data and security must be considered as critical requirements for a successful implementation of AI in manufacturing [10, 11, 13, 18, 19]. Therefore, we determine them as the third and cross-sectional domain in our conceptual framework (see Fig. 1).

2.2 Dimensions of the Model

Guided by the fact that the boundaries between the tangible and intangible innovations in the context of digital technologies used in manufacturing are difficult to define, we decide not to consider them separately, but to integrate them in two perspectives. We therefore combine smart products and smart services in one dimension, the *digital product-service portfolio*. Additionally, we develop the dimension *digital production* by combining smart factory and smart operations in the process perspective. Moreover, due to their high-significance for AI, we add *data and security* as a dimension to our model. This understanding results in an AI-readiness model

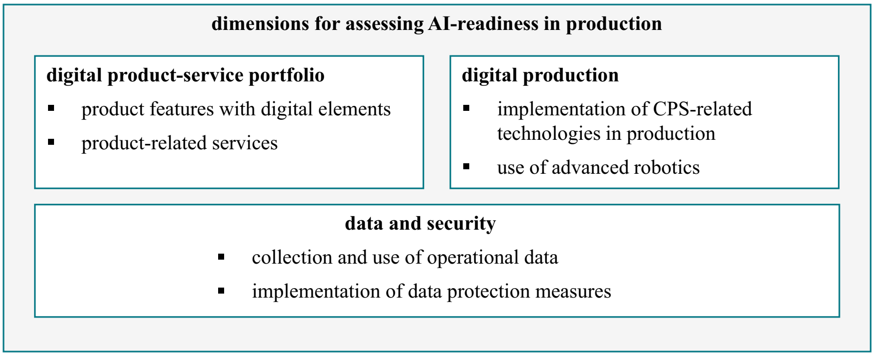


Fig. 2 Three dimensions for assessing AI-readiness in production

with three dimensions: digital product-service portfolio, digital production, and data and security. As we consider the latter as the foundation for the other two dimensions, we define it as the first dimension. Each dimension can be broken down into individual components as well as further differentiated and operationalized using indicators in order to measure AI-readiness in practice. Figure 2 visualizes the three dimensions with their respective sub-areas considered in our model.

In the first dimension, data and security, we analyze the collection and secure handling of data. We thereby consider aspects of data availability through automated operational data storage as well as the implementation of various information technology (IT) security measures for the protection of this data. Data security is considered a prerequisite for ensuring protected data handling [18]. As this dimension is significant for both the product and production domains, we classify it as the basis for the first two dimensions.

In the second dimension, the product-service portfolio, we research the company's product mix and its digital features. In addition to products equipped with digital elements, we also consider digital services accompanying the products. Thus, the focus is the extent to which information and communications technology (ICT) components enable products to collect and communicate data and whether the firm offers services (e.g. condition monitoring, predictive maintenance) based on the smart components.

Our third dimension, digital production, considers production processes and examines the use of specific technologies and applications driving digitization in the production environment. These are an expression of a profound technological digitization in the production setting and thus open up access to the opportunities associated with highly automated production in a digitized factory [16].

2.3 Assessment of AI-Readiness

In order to map the AI-readiness of manufacturing companies, we developed a metric index that includes the readiness scores for each dimension, with a maximum of two points per dimension. Our scoring procedure is based on the following logic.

In the first dimension (data and security), the scoring depends on the interplay between the different sub-categories. As we consider data and IT security as essential foundations for AI [18, 19] and therefore also for the AI-readiness, we assume that companies must have taken initial steps in order to receive one point. Only those companies implementing both far-reaching IT security measures (such as hardware and software security solutions in combination with specific trainings) and storing their operational data automatically score full points.

In the second dimension (digital product-service portfolio), the sub-dimensions also influence the allocation of points. A manufacturing company achieves one point if its product is equipped with at least one digital technique used for e.g. collecting data through sensors or for controlling some mechanisms or systems of the product. A firm only achieves the full score if it additionally offers a digital service (e.g. predictive maintenance or condition monitoring).

For the third dimension (digital production), we consider two different technology areas in production, the use of which we assume to be a prerequisite for AI implementation. If the manufacturing firm applies either basic cyber-physical system (CPS)-related technologies (e.g. advanced manufacturing execution systems (MES)) or robotic technologies in production, it scores one point. Only those companies using a variety of wide-reaching digital technologies in different areas of production receive the maximum of two points.

Combining the three dimensions, the index value for AI-readiness can range from a level of zero (no AI-readiness) to the highest level with a maximum of six points (strong AI-readiness). Based on this logic, we distinguish four groups of AI-readiness: No readiness, low readiness, medium readiness and high readiness (see Fig. 3).

The index score of the lowest level on our scale corresponds to the group of companies with no readiness for AI. These companies neither offer a digital product service portfolio, nor use digital technologies in production. Similarly, they have not yet implemented any measures in terms of data collection and IT security. Manufacturing companies in the low AI-readiness group have already implemented some initial measures and/or adopted some basic digital techniques. However, in at least one of the three considered dimensions, they have not implemented any technologies or relevant measures. The group of companies with a moderate level of AI-readiness has already created supportive foundations in at least two of the three dimensions. Thus, initial approaches for expanding the technical foundations on the path to AI-readiness in production are discernible in this group of companies. Finally, a digital product-service portfolio as well as the use of digital production technologies characterize the top group with high AI-readiness. In this group, high AI-readiness runs through all three dimensions including the systematic collection

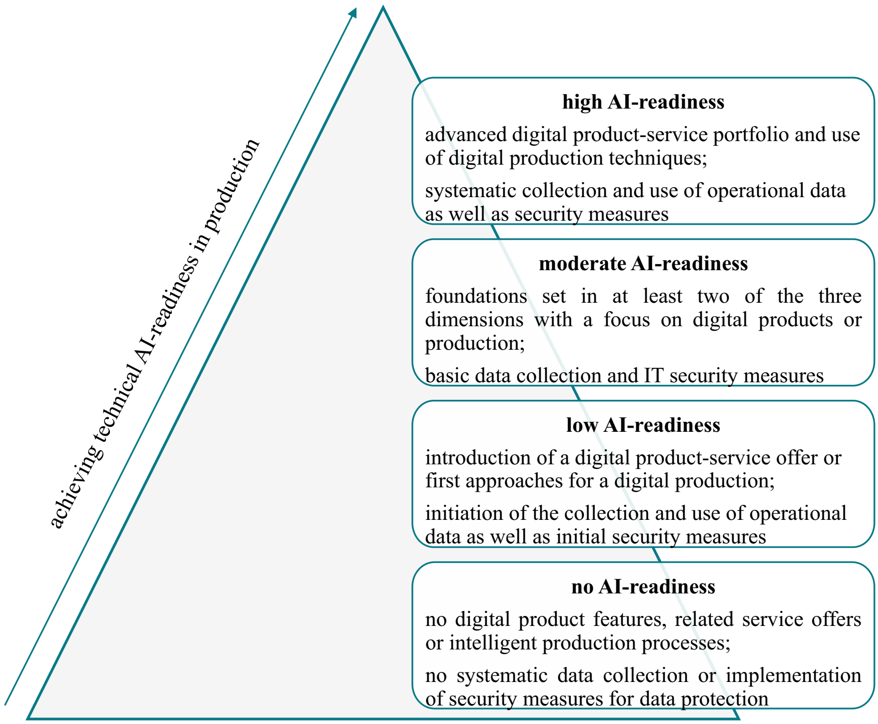


Fig. 3 Levels for achieving AI-readiness in production

and use of operational data as well as the implementation of comprehensive IT security measures.

3 Illustrative Case Studies

3.1 Case ALPHA—Product Specialist

ALPHA, a medium-sized company (approx. 60 employees), focuses on special machine construction and individually engineers solutions for its customers in the automotive and aerospace industries. At the core of the company is the construction of lifting equipment and assembly systems as well as automation solutions, such as robotic systems. ALPHA purchases the required parts and assembles individual solutions for its customers, drawing on its know-how in different areas (production processes as well as software development).

Looking at our AI-readiness index, ALPHA shows a clear focus on the digital product-service portfolio, supported by a basic data and IT security foundation, while a digital production is less prominent. The company successfully integrates different digital elements in its innovative products, such as RFID tags, various interfaces as well as sensor technologies. Based on these intelligent elements, ALPHA offers product-related services, e.g. digital remote monitoring. Since ALPHA does not manufacture standardized machines but rather designs custom solutions, many production processes lack automation. Although some CPS-related technologies are applied in assembly (e.g. through a warehouse management system), there is no extensive use of digital technologies in production. ALPHA collects some operational data, but most of the engineering knowledge remains in the heads of the designers and assembly workers. The company attaches great importance to IT security and sensitizes its employees accordingly. In addition, a few software solutions are in place to protect data.

To sum up, the evaluation of AI-readiness according to our index leads to the classification of ALPHA in the group of moderate AI-readiness with an overall index score of three points. The focus of the company is on a highly digital product-service portfolio with a moderate data and security basis.

3.2 Case BETA—Production Specialist

The medium-sized production company BETA employs approximately 130 people and specializes in turning and milling of precision parts. It manufactures a variety of different metal parts in medium to large series for its customers in automotive, aerospace and medical technology sectors. The digitalization of production plays an essential role in the company and is constantly being driven forward. Nevertheless, the machine park also includes some older machines that are not yet equipped with the latest technologies.

When analyzing the dimensions of the AI-readiness index, BETA reveals a clear focus on digital production as well as data and security, while a digital product-service portfolio is somewhat neglected. BETA's products are not equipped with any digital elements or accompanied by product-related services. One of the reasons for this is that BETA produces metal parts that its customers subsequently process. In contrast, BETA's digital production is already highly advanced. The company uses CPS-related technologies in production, including a production control system and warehouse management system. In addition, various robots support the production of large series and collaborate with the workers. BETA's IT security and data handling is also advanced. The manufacturing firm stores operating data automatically enabling various analyses of the production process. To ensure the protection of the sensitive operational data, the company trains its employees regularly. Various barriers, such as authentication measures, restrict access to the production hall. Hardware and software solutions provide additional IT security.

According to our AI-readiness index, BETA thus records moderate AI-readiness (total index score: four points), with strong readiness in all dimensions except the digital product-service portfolio.

4 Concluding Remarks

Based on the existing literature, we developed a conceptual framework of a manufacturing-specific model that allows us to assess the technological AI-readiness of manufacturing companies in production. We visualized this assessment using a qualitative analysis in two use cases. These illustrative presentations reinforce our understanding of the importance of possible clusters on AI-readiness. The manufacturing companies analyzed focus on either digital production or a digital product, while data and security form an essential foundation for both priorities.

To continue developing the AI-readiness model for manufacturing, we will operationalize the categories as well as sub-categories into a set of indicators and conduct a quantitative analysis of AI-readiness in the manufacturing sector using a large-scale survey. In this process, we plan to explore the clusters of companies (e.g. product experts or production experts) further as well as research differences in production performance. Since the use of AI in production areas promises to increase performance (esp. with respect to productivity and efficiency [5]), this quantitative analysis will also evaluate statements about the economic impact of AI-readiness. In addition to the quantitative analysis, we plan to conduct an in-depth qualitative research for examining reasons or factors that lead to success based on a specific combination of technologies, measures or individual clusters. These analyses aim to support decision makers in leveraging the results on AI-readiness to successfully integrate potential applications into production processes.

References

1. Ritter, T., Pedersen, C.L.: Digitization capability and the digitalization of business models in business-to-business firms: past, present, and future. Ind. Mark. Manage. **86**, 180–190 (2020). https://doi.org/10.1016/j.indmarman.2019.11.019
2. Benner, M.J., Waldfogel, J.: Changing the channel: digitization and the rise of “middle tail” strategies. Strateg. Manag. J. **86**(7), 1–24 (2020). https://doi.org/10.1002/smj.3130
3. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2019). https://doi.org/10.1016/j.artint.2018.07.007
4. Mikalef, P., Fjørtoft, S.O., Torvatn, H.Y.: Developing an artificial intelligence capability: a theoretical framework for business value. In: Abramowicz, W., Corchuelo, R. (eds.) Business Information Systems Workshops, Bd 373, pp. 409–416. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-36691-9_34
5. Sanchez, M., Exposito, E., Aguilar, J.: Autonomic computing in manufacturing process coordination in industry 4.0 context. J. Ind. Inf. Integr. **19**(100159) (2020). https://doi.org/10.1016/j.jii.2020.100159
6. Horvat, D., Stahlecker, T., Zenker, A., Lerch, C., Mladineo, M.: A conceptual approach to analysing manufacturing companies’ profiles concerning industry 4.0 in emerging economies. Procedia Manuf. **17**, 419–426 (2018). https://doi.org/10.1016/j.promfg.2018.10.065
7. Schumacher, A., Erol, S., Sihn, W.: A maturity model for assessing industry 4.0 readiness and maturity of manufacturing enterprises. Procedia CIRP **52**, 161–166 (2016). https://doi.org/10.1016/j.procir.2016.07.040

8. Neugebauer, R., Hippmann, S., Leis, M., Landherr, M.: Industrie 4.0—from the perspective of applied research. Procedia CIRP **57**, 2–7 (2016). https://doi.org/10.1016/j.procir.2016.11.002
9. Gökalp, E., Şener, U., Eren, P.E.: Development of an assessment model for industry 4.0: industry 4.0-MM. In: Mas, A., Mesquida, A., O'Connor, R.V., Rout, T., Dorling, A. (Hrsg.) Software Process Improvement and Capability Determination. 17th International Conference, SPICE 2017, Palma de Mallorca, Spain, 4–5 Oct 2017, Proceedings, Bd 770, pp. 128–142. Springer, Cham (2017)
10. Eljasik-Swoboda, T., Rathgeber, C., Hasenauer, R.: Artificial intelligence for innovation readiness assessment. In: The 2nd Annual International Symposium on Innovation and Entrepreneurship of the IEEE Technology and Engineering Management Society, Zhejiang University, Hangzhou, China, 24–26 Oct 2019, S 1–6. IEEE, Piscataway, NJ (2019)
11. Ellefsen, A.P.T., Oleśków-Szłapka, J., Pawłowski, G., Toboła, A.: Striving for excellence in AI implementation: AI maturity model framework and preliminary research results. Logforum **15**(3), 363–376 (2019). https://doi.org/10.17270/J.LOG.2019.354
12. Holmstrom, J.: From AI to digital transformation: the AI readiness framework. Bus. Horiz. In Press, Journal Pre-Proof (2021). https://doi.org/10.1016/j.bushor.2021.03.006
13. Jöhnk, J., Weißert, M., Wyrtki, K.: Ready or not, AI comes—an interview study of organizational AI readiness factors. Bus. Inf. Syst. Eng. **63**(1), 5–20 (2021). https://doi.org/10.1007/s12599-020-00676-7
14. Akdil, K.Y., Ustundag, A., Cevikcan, E.: Maturity and readiness model for industry 4.0 strategy. In: Ustundag, A., Cevikcan, E. (eds.) Industry 4.0, pp. 61–94. Springer, Cham (2017)
15. Kirner, E., Kinkel, S., Jaeger, A.: Innovation paths and the innovation performance of low-technology firms—an empirical analysis of German industry. Res. Policy **38**(3), 447–458 (2009). https://doi.org/10.1016/j.respol.2008.10.011
16. Lichtblau, K., Stich, V., Bertenrath, R., Blum, M., Bleider, M., Millack, A., Schmitt, K., Schmitz, E., Schröter, M.: Industrie 4.0-Readiness. Aachen, Köln (2015)
17. Frietsch, R., Koschatzky, K., Beckert, B., Jackwerth, T., Lerch, C., Lubczyk, M., Neuhäusler, P., Rammer, C., Stahlecker, T., Viete, S.: Struktur und Dynamik des Forschungs- und Wissenschaftsstandorts Baden-Württemberg. Schriftenreihe der Baden-Württemberg Stiftung (2020)
18. Baryannis, G., Validi, S., Dani, S., Antoniou, G.: Supply chain risk management and artificial intelligence: state of the art and future research directions. Int. J. Prod. Res. **57**(7), 2179–2202 (2019). https://doi.org/10.1080/00207543.2018.1530476
19. Dogru, A.K., Keskin, B.B.: AI in operations management: applications, challenges and opportunities. J. Data Inf. Manag. **2**(2), 67–74 (2020). https://doi.org/10.1007/s42488-020-00023-1

Architecture for Predictive Maintenance Based on Integrated Models, Methods and Technologies

Andreas Werner, Roi Mendez-Rial, Pablo Salvo, Vasiliki Charisi, Joaquín Piccini, Alireza Mousavi, Claudio Civardi, Nikos Monios, Diego Bartolomé Espinosa, Marlène Hildebrand, Nikolas Zimmermann, Irati Vizcarguenaga Aguirre, Jacopo Cassina, Diego Nieves Avendano, Helder Oliveira, Daniel Caljouw, Matteo Fazziani, and Silvia de la Maza

Abstract The evolution to Industry 4.0 is creating the impetus for the manufacturing industry to increase productivity through smart management and stabilization of resources, capacity and utilisation. Increased plant availability, extended service life of resources as well as optimised product and process quality require

A. Werner (✉)
University of Stuttgart IAT, Institute of Human Factors and Technology Management, Nobelstrasse 12, 70569 Stuttgart, Germany
e-mail: andreas.werner@iat.uni-stuttgart.de

R. Mendez-Rial
AIMEN Technology Centre, Calle Relva Torneiros 27 A, 36410 Porrino, Spain

P. Salvo
Asociacion de Empresas Tecnologicas Innovalia, Calle Rodriguez Arias, 6 DPTO 605, 48008 Bilbao, Spain

V. Charisi
ATLANTIS Engineering SA, Antoni Tritsi 21, 55102 Thessaloniki, Greece

J. Piccini
Autotech Engineering, AIE, Parque Empresarial Boroa P2A 4 Local, 48340 Bizkaia, Spain

A. Mousavi
Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK

C. Civardi
CDS S.R.L., Via del Commercio 91/97, 29122 Piacenza, Italy

N. Monios
CORE Innovation and Technology OE, Delagrammatika 5, 34100 Chalkis, Greece

D. B. Espinosa
CRIT Srl, Via Confine 2310, 41058 Vignola, Italy

M. Hildebrand
Ecole Polytechnique Federale de Lausanne, Route Cantonale, 1015 Lausanne, Switzerland

N. Zimmermann
Fraunhofer Institute for Industrial Engineering IAO, Nobelstrasse 12, 70569 Stuttgart, Germany

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_25

intelligent maintenance strategies. The conventional reactive maintenance (run-to-failure) causes unexpected production stoppages, and preventive maintenance at times leads to waste of working hours and material due to the premature replacement of machine components. A smart Predictive Maintenance (PdM) strategy equipped with fault detection and prediction based on acquired, processed and analysed data can result in an accurate estimation of the Remaining Useful Life (RUL) of machine components and thus trigger appropriate maintenance action plans. Data acquisition, processing, analysis and rule-based decision supporting require the development, application and combination of various Industrial Internet of Things (IIoT) devices, models and methods in an integrated manner. Through transparent development and integrated harmonisation of all models, methods and technologies, fault detections and respective RUL estimations of machine components become more accurate and reliable. This leads to an increasing acceptance of employees towards software-based recommendations, in particular maintenance instructions for operators and proposals for an optimised development of the next generation of production systems and equipment. Within the scope of the EU-funded project Z-BRE4K, this paper proposes an IIoT architecture that presents models, methods and technologies in an integrated manner and highlights the data and information flow between them. The architecture including the infrastructure has been applied in three pilot cases with the industrial end users PHILIPS, GESTAMP and CDS to demonstrate the compatibility of the architecture to different industries with various production systems and diverse conditions, requirements and needs. Based on the adaption of the generic architecture for the pilot cases, the models, methods and technologies were developed efficiently and continuously improved and validated. The proposed

I. V. Aguirre
Fundación AIC-Automotive Intelligence Center Fundazioa, Parque Empresarial Boroa P2A-4, 48340 Amorebieta-Etxano, Spain

J. Cassina
HOLONIX S.R.L., Corso Italia 8, 20821 Meda, MB, Italy

D. N. Avendano
imec—IDLab at Ghent University, Technologiepark-Zwijnaarde 126, 9052 Ghent, Belgium

H. Oliveira
Inovamais—Servicos de Consultadoria em Inovacao Tecnologica S.A., Rua Dr Afonso Cordeiro 567, 4450-309 Matosinhos, Portugal

D. Caljouw
PHILIPS Electronics Nederland B.V., Boschdijk 525, 5621 JG Eindhoven, The Netherlands

M. Fazziani
SACMI Cooperativa Meccanici Imola Societa Cooperativa, Via Selice Provinciale 17/A, 40026 Imola, Italy

S. de la Maza
Trimek SA, Camino de la Yesera 2, 01139 Zuya Alava, Spain

architecture is intended to be applicable across industries to facilitate the transformation from reactive or preventive to PdM and thereby improve the competitiveness of manufacturing companies.

Keywords Predictive maintenance · Industrial Internet of Things architecture · Fault detection · Remaining Useful Life estimation

1 Introduction and State of the Art

In order to advance the competitiveness of manufacturing companies, their productivity must be increased, for which smart maintenance strategies are required. Predictive Maintenance (PdM) enables an increased availability, service life and performance of plant and equipment as well as an optimised product and process quality. Through a comprehensive analysis of production systems, the Remaining Useful Life (RUL) of machine components can be estimated and thus the appropriate point in time for maintenance can be determined [1]. This leads to a higher Overall Equipment Effectiveness and thus to an increased productivity. While reactive maintenance only provides for repairs after a breakdown, preventive maintenance aims at replacing components at predefined time intervals based on manufacturer specifications and empirical values [2]. Condition-based maintenance includes continuous condition monitoring and the resolution of maintenance measures based on defined condition deviations and events. Currently, most companies rely on reactive or preventive maintenance strategies, as costs and complexity of more proactive maintenance strategies initially discourage. Especially PdM poses challenges for many companies, which is confirmed in a study from 2017 [3].

Existing approaches in literature mostly focus on applying models, methods and technologies on specific use cases [4, 5] or highlight the most common techniques in the context of PdM, e.g. Long Short-Term Memory (LSTM)—according to Wu et al. [6] commonly used due to its efficiency in processing time series data—and linear regression [7, 8], or mechanisms for feature extraction as an important pre-processing step for training the models [9]. Furthermore, multiple literature deals with hybrid modelling approaches, e.g. combining various machine learning algorithms [10], data-driven and physics-based models [11, 12] or data mining and ontology-based semantics [13]. Even though scattered approaches exist regarding PdM strategies from an architecture and design to its practical implementation [14], companies still cite a lack of a systematic approach—in the sense of a clear procedure based on an Industrial Internet of Things (IIoT) architecture to derive implementation steps—as a crucial obstacle for establishing a corresponding business model and thus for implementing PdM into their manufacturing and process settings [3], which represents a research gap.

To close this research gap, this paper presents an approach for a generic IIoT architecture based on integrated models, methods and technologies with the focus on providing a basis for implementing PdM into manufacturing companies from different industries with various production systems and diverse conditions, requirements and needs. The proposed architecture combines models, methods and technologies within Industry 4.0 compliant layers such as data acquisition, transfer and persistence, analysis and business logic. By implementing PdM at PHILIPS, CDS—together with the original equipment manufacturer SACMI—and GESTAMP within the scope of the EU-funded project Z-BRE4K, the applicability of the generic architecture and the included models, methods and technologies is demonstrated. For the development of the architecture as well as for the included models, methods and technologies, established approaches are used as an orientation, but are continuously advanced and combined according to the conditions, requirements and needs of the three pilot cases. Particular attention is paid to the transparent development of the models, methods and technologies with continuous consultation with the industrial end users and their employees in order to increase acceptance towards the applications. The overall result is a holistic architecture for PdM that, subject to minor adjustments, is applicable to a wide range of industries and use cases of manufacturing companies.

The structure of the paper is as follows: First, the proposed generic architecture for PdM, highlighting its layers is presented. In Sect. 3, the application of the general architecture for the three pilot cases is described, however it only gives a short overview with exemplary excerpts of the three use cases to demonstrate the compliance of the generic architecture to different industries with various production systems and diverse conditions, requirements and needs. Finally, the paper is summed up and an outlook is given.

2 Approach for a Generic Architecture for Predictive Maintenance

Figure 1 presents an approach for a generic holistic architecture as a basis for the implementation of PdM in manufacturing and process settings subject to minor customization. Its successive layers of data acquisition, transfer and persistence and analysis, as well as the layer in which decisions are made based on generated information, the business logic, will be described in more detail in the following sections.

2.1 Data Acquisition

The layer of data acquisition focuses on the shop floor of a manufacturing company and includes the entire machinery and the measuring equipment, in particular sensors. Furthermore, this layer is source of expert knowledge about the machines and the production process, including engineering data, e.g. material specifications or data

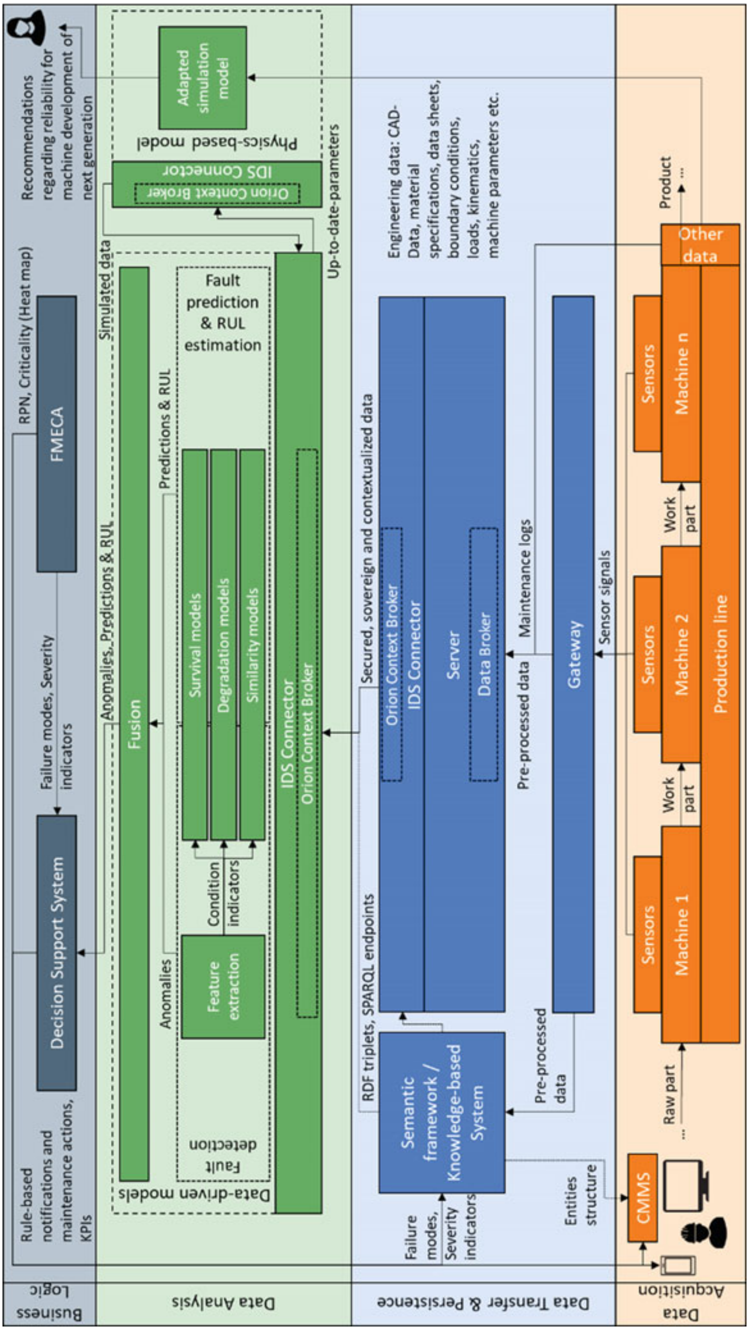


Fig. 1 Approach for a generic architecture for PdM

sheets of different components of the production systems. Measuring principles and sensor types must be selected or retrofitted depending on the required signals. This selection process can be supported using existing toolboxes, e.g. according to Fleischer et al. [15]. Furthermore, the sensors must be installed and mechanically integrated at an appropriate location. The sensor readings are transmitted to gateways for pre-processing of the signals using appropriate communication technology, e.g. bus systems.

2.2 Data Transfer and Persistence

The acquired sensor data is processed via gateways with functionalities such as interoperability, aggregation and—if required—local data pre-processing. Interoperability simplifies the connection of multiple machines and other devices, using a variety of interfaces, protocols and standards for both local installation (edge) and remote data transmission (connectivity). An ontology-based semantic framework can be applied to foster interoperability, i.e. to maintain consistency across different data sources and sinks. It involves semantic modelling of maintenance entities (e.g. production systems, work parts, failure modes) and actions, which then serve as a reference model that allows linking all data in the respective use case and providing standards-based access to all data related to a particular application. By brokering data coming from a wide variety of publishing sources, the data is centralized and contextualized, which then can be subscribed by data analysing consumers. In order to ensure the data security and sovereignty of the transferred and persistent data, Industrial Data Space (IDS) aims to formulate an infrastructure for the fast, secure and sovereign transmission and use of data. Only certified participants whose identity has been verified beforehand are allowed to enter the data space. IDS connectors based on FIWARE specification use system adapters to receive data from various sources, e.g. from a Message Queuing Telemetry Transport (MQTT) broker built on top of the gateway, and convert them into a standardized format. Consumers are able to subscribe contextualized data from an integrated Orion Context Broker (OCB) in a standardized and secure manner via Next Generation Service Interfaces (NGSI), which is a protocol developed by Open Mobile Alliance (OMA) to manage context information [16, 17].

2.3 Data Analysis and Business Logic

The layer of data analysis aims to detect and predict faults and estimate the RUL of critical machine components. Considering a deterioration profile of a machine and its components, the RUL is the period from a current health status of the machine to a detected fault based on computed condition indicators. Fault detection includes the distinguishing between a healthy and a faulty status of a machine. Various reactive algorithms can be applied to detect anomalies within the data, tailored to the requirements of the respective use case. Some sensor readings do not show a significant changing trend between healthy states and faults and therefore do not contribute to the selection of useful features for training a model. Consequently, data reduction needs to be performed by selecting only sensor signals with the strongest trend and combining them to calculate condition indicators by means of feature extraction mechanisms.

There are three common approaches to predict faults and to estimate the RUL, depending on the data available [18]. First, if no data is available on the history of the running machine, but data about breakdowns as single events, then survival models, i.e. probability density functions, can be used. This approach does not involve data collection of machine behaviour prior to a failure. The second approach for RUL estimation is once complete run-to-failure data (data from healthy state, degradation and actual failure) of the machine in operation or similar machines are available, then similarity models should be used. For similarity models, a lot of training data is needed so that the machine in operation can be aligned with these historically most similar degradation profiles. In cases where failures cannot be detected and therefore no failure data is available to train a predictive model, (safety) thresholds must be defined for condition indicators that should not be crossed. In that case, a fed degradation model is useful, as the behaviour of the running machine is monitored and by identifying a pattern within the data, a trend can be extrapolated until the threshold is reached, e.g. via regression analysis. Another approach in cases where failures could not be detected due to missing machine failures is to simulate potential failures using physics-based models of the machine and its components. By the hybridization of data-driven and physics-based models, simulations can be carried out quickly and cost-effectively to expand the amount of data required. In addition, using physics-based models, physical causes and effects can be investigated to gain additional knowledge, potential failure mechanisms can be identified and faults can be localized more easily.

Regarding the business logic, a decision support system (DSS) evaluates the performance of the machines and receives anomalies, predictions and RUL estimations published by the data-driven models, combining them into a single result by automatically tuning and promoting the most effective predictions approach. The DSS decides on preventive actions by activating recommendations to improve maintainability and operational efficiency on the shop floor. A feedback loop by the operator can be applied, which takes into account their opinions about the recommendations and their quality. The Failure Mode and Effects Analysis (FMEA) identifies the potential failure modes, causes and effects associated with a machine component and how the performance of the system is affected, addressing each failure mode and its respective effects in the system independently. FMECA additionally analyses the criticalities of each potential failure mode and effect and calculates the Risk Priority Number (RPN) based on the combination of occurrence and severity of a given combination of failure mode, cause and effect. Based on that, a heat map provides information to the user in an intuitive manner. The user can immediately see whether a particular failure mode causes a high, medium or low risk to the system based on the area placed in the heat map.

3 Application of the Generic Architecture in Pilot Cases

The general architecture proposed in Sect. 2 is applied in three pilot cases with industrial end users to proof the compatibility of the architecture to different industries with various production systems and diverse conditions, requirements and needs. One pilot case deals with cold forming for the manufacturing of cutting elements for electric shavers at the PHILIPS factory. The second one is about continuous compression moulding for the manufacturing of plastic closures at CDS and the third one involves cold forming, arc-welding and in-line quality inspection for the manufacturing of chassis parts at GESTAMP.

3.1 Cold Forming for the Manufacturing of Shaver Cutters

Shaving systems are manufactured at the PHILIPS factory. One of the parts of a shaving system is a cutter. These cutters are produced on a production line consisting of cold forming, finishing, measuring and assembly. Figure 2 presents the implementation of the generic architecture to this pilot case.

Data acquisition. This pilot case focuses on dies of a press that form the cutters from a metal strip entering the press. Multiple acoustic emission (AE) sensors are installed in crucial positions that can measure signals in terms of changes in the hardened steel.

Data transfer and persistence. The AE data along with quality inspection information are collected in real-time by the measuring equipment coupled to the factory network. After pre-processing, a gateway sends the data to a database and publishes them to a MQTT broker extended with an IDS connector. Via the counterpart of the IDS connector provider, the IDS connector consumer, the data contextualized by an integrated OCB are transferred to the data-driven models to detect and predict faults. The semantic framework is based on an ontology, structuring the entities of the IDS connectors. The cold forming dies and all related parts, embedded data sources, failure modes, severity indicators and maintenance actions are modelled as an enumeration of related entities to represent the pilot case-specific knowledge machine-understandably.

Data analysis and business logic. By subscribing the AE data as well as other data sources, e.g. maintenance logs, in JSON format from the OCB using NGSI API, various algorithms are applied to detect and predict faults of the cold forming dies. To detect faults, i.e. AE waves that do not conform to neighbouring signals, a simple

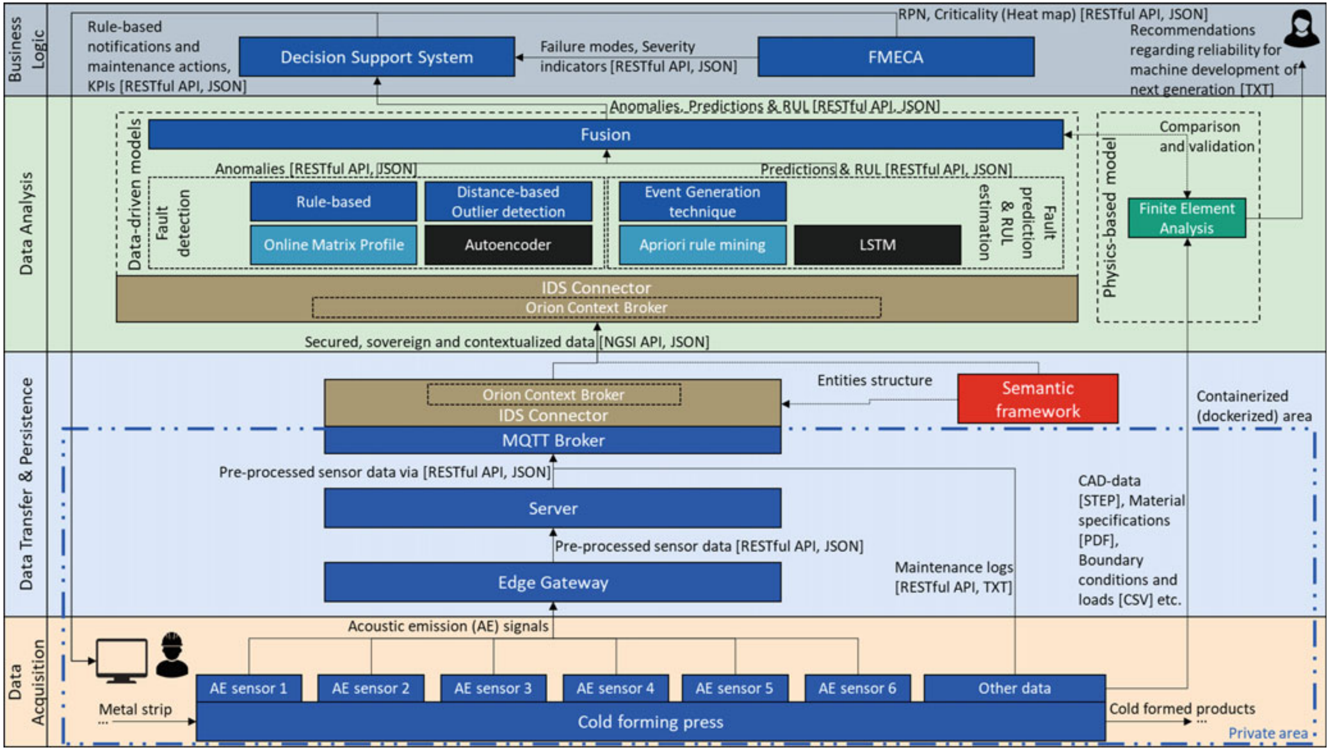


Fig. 2 Adapted architecture for PdM at PHILIPS

rule-based approach monitors specific sensorial inputs to detect rule violations based on user-defined thresholds. Since the threshold specification is difficult due to a lot of instability to the sensorial input, a distance-based outlier detection approach is used, which alleviates the need for rule specifications for each different sensorial input. This approach is based on the Micro-cluster Continuous Outlier Detection algorithm and can be applied to streaming data. For fault prediction an event-based algorithm is utilized, identifying patterns of events that occur before failures, in order to train the models for predictions. An approach for discretization of the input signal is then used in order to convert the sensorial signal to artificial events that might or might not have an actual meaning to the physical world. It should be noted that a number of other algorithms (e.g. Autoencoders, matrix profiles, Apriori) are used for both fault detection and prediction for validation purposes and thus to increase the reliability of calculations. Besides the aforementioned data-driven models, the physics-based model of the dies is represented as a Finite Element Method (FEM) model, analysing critical parts against high cycle fatigue to derive recommendations to strengthen the weakest parts, supporting employees during the design phase of the next generation of machine components to optimize the reliability of the machine and thereby minimize breakdowns.

In order to combine multiple fault detections or predictions, a fusion component is applied to generate higher-level detection and prediction strategies, e.g. the combination of monitoring the first and the second die of the press and only if faults are detected in both dies, a report is generated. This leads to a reduction of false positive reports. Finally, a DSS receives input from both a FMECA, mainly failure modes and severity indicators, as well from the fusion component, and uses a rule engine, which is able to decide automatically upon specific actions to mitigate the detected and predicted faults. The DSS applies business related rules that are created by the shop floor managers based on their experiences and knowledge.

3.2 Continuous Compression Moulding for the Manufacturing of Plastic Closures

Plastic closures for the food and beverage industry are manufactured at CDS. This paper particularly addresses the production of plastic closures carried out by means of compression moulding techniques. In this regard, CDS produces various formats of products with many diverse parameters such as material, dimensions or weight. Figure 3 presents the implementation of the generic architecture to this pilot case.

Data acquisition. The focus of this pilot case are five Continuous Compression Moulding (CCM) machines—designed and manufactured by SACMI—with three auxiliary modules, a hydraulic unit (HU), a plastic extruder (EX) and a thermal regulator (TH) equipped with several sensors to measure pressures, temperatures and volume flow rates. The TH provides an aqueous fluid to cool the cavities of the CCM machines and will be considered in more detail in the following paragraphs.

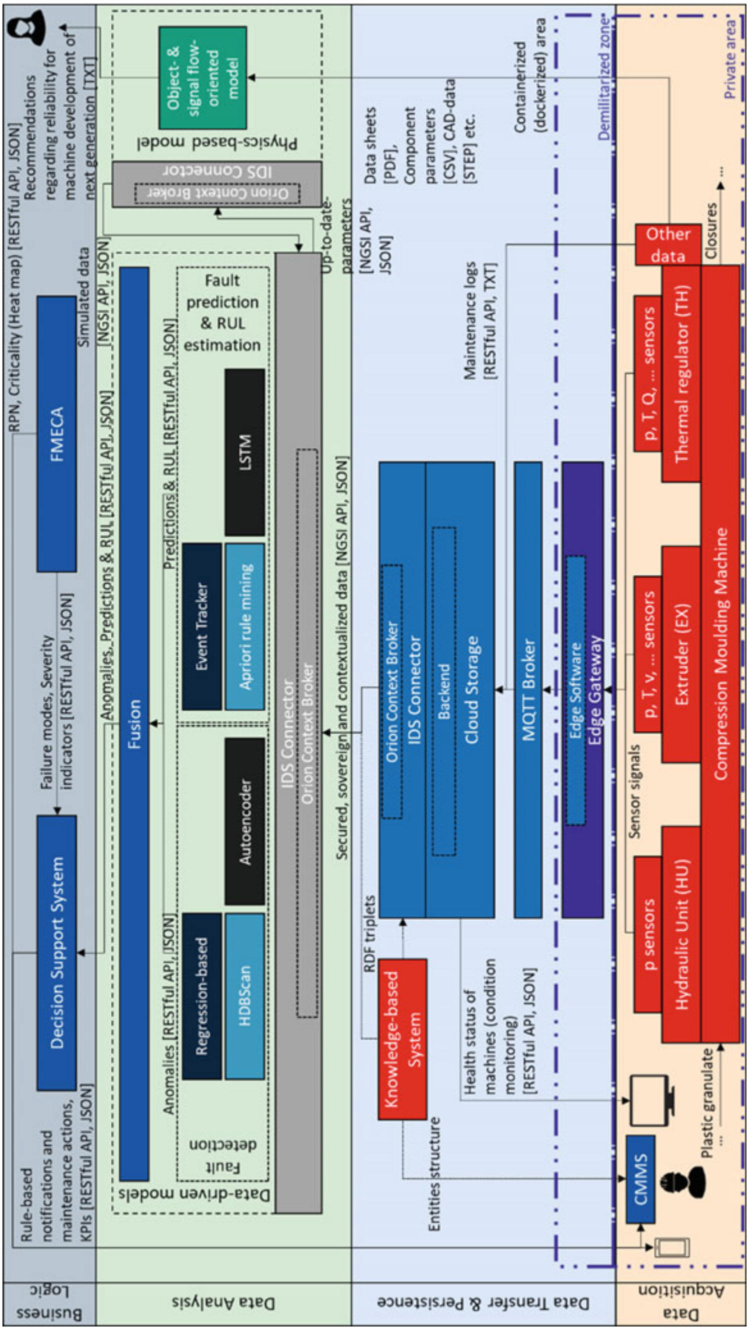


Fig. 3 Adapted architecture for PdM at CDS

Data transfer and persistence. The sensor signals are collected in a gateway. To make the data available to the subsequent IIoT infrastructure, additional software is installed on the gateway. This allows publishing the data to a MQTT broker, enabling a cloud storage as intermediate layer to publish data to an integrated backend, generating a first data visualization to provide early information to the operator about the status of the machine. All data is transformed in Resource Description Framework (RDF) triplets and stored in a triple store database. On top of the database, a reasoner supports the extraction of knowledge according the consumer's queries, building a knowledge-based system. The cloud storage is extended with an IDS connector to ensure data security and sovereignty, following the same policy as in Sect. 3.1.

Data analysis and business logic. By subscribing the sensor data in JSON format from an OCB using NGSI API, various algorithms are applied to detect faults and to estimate the RUL. For instance, a regression-based approach is used to detect faults of the TH—which are mainly caused by a clogging filter—based on the trend of incoming sensor data. The approach uses historical data and detects violations based on predefined thresholds that are either linear or exponential to the measured trend. By applying an event clustering algorithm, the extrapolated trends of the individual sensor values based on the regression analysis are weighted up to the thresholds, so that an overall RUL of the filter is calculated. Due to missing machine breakdowns and thus a lack of detected faults, a physics-based representation of the TH as an object- and signal flow-oriented modelling approach with integrated virtual sensors is applied for multi-physics simulations of the TH. The hybridization of the regressive event tracker model and the physics-based model is carried out by varying various parameters in the physics-based simulation model, automatically integrating virtual sensor values into the regressive event tracker model, while at the same time changing parameters during operation are integrated into the physics-based model.

Reports based on a fusion algorithm are triggered when pre-defined combinations of outcomes of the models are reached. Besides the rule-based processing of these outcomes, the DSS requests information from a FMECA tool, in particular the level of criticality for failure modes and checks the combination of the confidence level along with the criticality and the timeframe of the predictions to activate maintenance actions, e.g. as a notification to a mobile device of the operator. Furthermore, a Computerized Maintenance Management System (CMMS) is implemented, so the operator can schedule maintenance actions and to interface with other entities such as the spare part warehouse inventory for an availability check.

3.3 Cold Forming and Arc-Welding for the Manufacturing of Chassis Parts

Lightweight chassis parts are manufactured at the GESTAMP facility. The production line includes a stamping cell with a servo-driven press for cold forming (bending and cutting) of the incoming steel sheets, a robot for arc-welding the formed parts and an in-line multi-sensor quality control system to ensure the quality of the finished manufactured parts. Figure 4 presents the implementation of the generic architecture to this pilot case.

Data acquisition. The stamping press is a closed system equipped with several pre-installed sensors for measuring torque, temperature, pressure and lubrication and two strain gage sensors installed at two connecting rods of the press. These locations are selected to provide the most significant strain history. The distribution of the load causes inertial forces that generate cyclic axial forces and stress, bending moments and stress perpendicular and parallel to the eccentric shaft. The tonnage signature provides important information that allows statements about the load, change in stock thickness and hardness, part lubrication, tool wear, stuck scrap in the die and the quality of the stamped parts. **The arc**-welding cell consists of two welding robots, a welding gun with a contact tip and an infrared (IR) sensor for an in-line quality control of the welding process. IR imaging provides information of the melt pool and surrounding areas during the welding process, such as geometry and temperature distribution. The IR system is comprised by a thermal camera and an embedded processing unit to perform the vision tasks in real-time and to ensure interoperability with subsequent applications.

Data transfer and persistence. All raw data acquired from each stroke of the stamping press as well as features extracted from the video sequence of the IR system together with other process parameters such as voltage and current are transferred as a XML or CSV file to a shared server. An IDS connector periodically queries this server via a system adapter. Once a new file is detected, the system adapter parses the information to NGSI. By sending the data to an integrated OCB, consumers are thus able to subscribe to a certain type of information in a standardized, secured and sovereign manner. The semantic framework is based on an ontology, following the same policy as mentioned in Sect. 3.1. Once the data is contextualized, consumers are able to analyse the data.

Data analysis and business logic. With regard to the stamping process, a perceptual metric known from quantifying image quality degradation caused by image processing, the Structural Similarity Index Measure, is used. The outliers are identified and the rate of failure determines the depreciation of the stamping dies. The instantaneous stamp force outside tolerance provides a defective part. In this regard, the defects rate provides the depreciation of the cold forming dies and therefore the RUL can be estimated. Due to only minor differences to the pilot case in Sect. 3.1 regarding the physics-based representation of the press as a FEM model, this will not be explained in more detail at this point.

Regarding the welding station, a video processing pipeline for condition monitoring and quality control combining edge and cloud processing is devised. In the edge, a feature extraction module for reducing the dimensionality of the data send to the cloud and a quality control classifier are implemented. First, a detection algorithm is applied to recognize and crop the region of interest around the contact tip during welding. Then, a bivariate classifier is combining convolutional (CNN) and recurrent

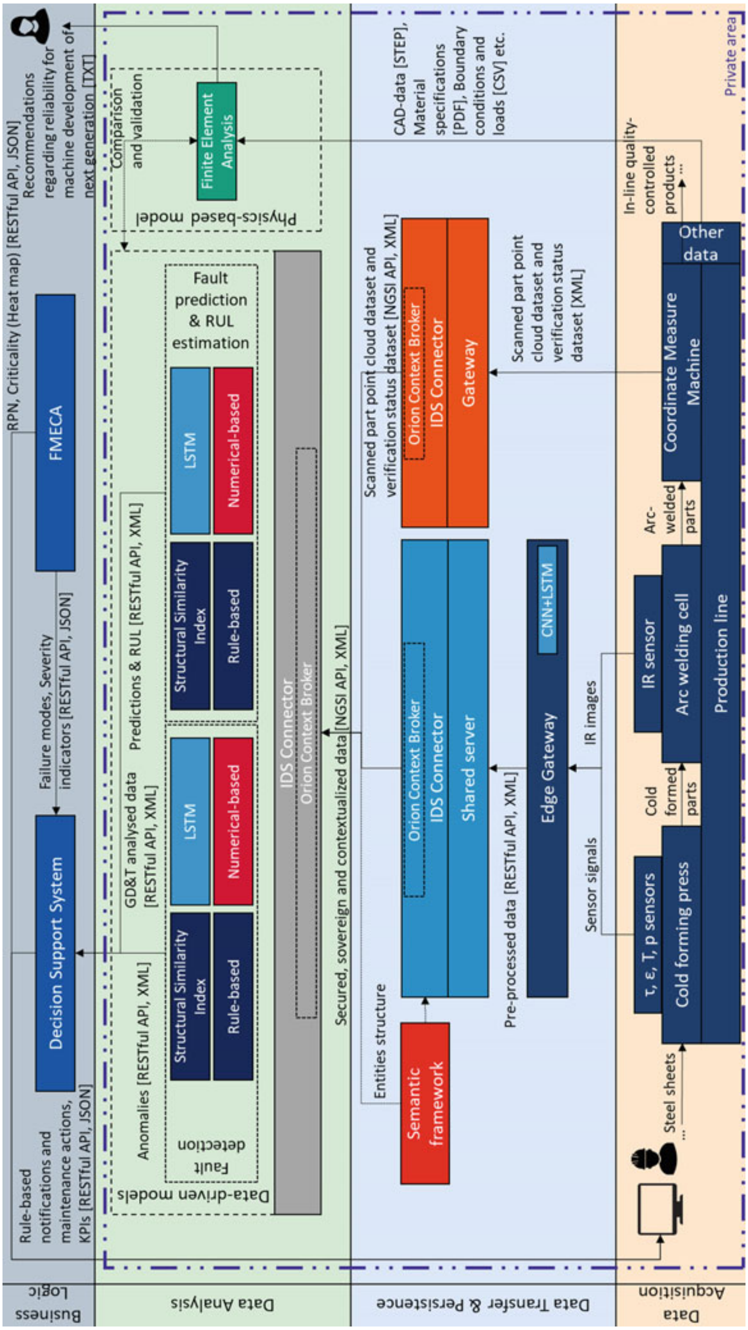


Fig. 4 Adapted architecture for PdM at GESTAMP

neural networks (LSTM) for a quality assessment to classify parts in defect and non-defect ones. The CNN serves for the extraction of spatial features, the LSTM-based network for the extraction of temporal features and a fully connected layer classifies the spatiotemporal features extracted. In the cloud, a single LSTM model driven by the extracted spatial features is applied in terms of condition monitoring. Finally, a geometric dimensioning and tolerancing (GD&T) analysis is applied to control the quality of the finished parts, using a point cloud of scanned parts to evaluate deviations by comparing the scan to a theoretical CAD model. The fusing of all data analysis outcomes is done in the context of the DSS, which follows the same logic (including the interaction with FMECA) as described in the two pilot cases above.

4 Summary and Outlook

In terms of the EU-funded project Z-BRE4K, an architecture for PdM based on integrated models, methods and technologies has been developed with the aim of implementing PdM into manufacturing companies. The proposed architecture including various models, methods and technologies, which were continuously advanced based on existing approaches according to requirements of the use cases, has been effectively applied at three companies due to its holistic nature and modularity. Thus, a successful transformation from reactive or preventive to PdM strategies at PHILIPS, CDS and GESTAMP has been realised. Due to the transparent development of the concepts and solutions, they are applied by employees of the companies in their daily work. The applicability of the architecture has been demonstrated and can now serve as a reference for companies from different industries with various production systems and diverse conditions, requirements and needs for implementing PdM.

In follow-up activities, measurements of improved key performance indicators and a profitability analysis should be carried out, as companies want to achieve a return on investment as soon as possible. Furthermore, the integration into higher-level systems must be expanded in order to establish PdM in the entire system landscape of a company.

Acknowledgements The content of this paper is part of the work in the project Z-BRE4K, which has received funding from the Horizon 2020 Framework Programme of the European Union under grant agreement no. 768869. The listing of co-authors is in alphabetical order according to the affiliation.

References

1. Mikat, H.: Hybride Fehlerprognose zur Unterstützung prädiktiver Instandhaltungskonzepte in der Luftfahrt. Technische Universität Darmstadt, Institut für Flugsysteme und Regelungstechnik, Dissertation (2015)

2. Werner, A., Zimmermann, N., Lentes, J.: Approach for a holistic predictive maintenance strategy by incorporating a digital twin. Procedia Manuf. **39**, 1743–1751 (2019). https://doi.org/10.1016/j.promfg.2020.01.265
3. Roland Berger GmbH: Predicitve Maintenance: Service der Zukunft—und wo er wirklich steht. Munich (2017)
4. Cubillo, A., Perinpanayagam, S., Esperon-Miguez, M.: A review of physics-based models in prognostics: application to gears and bearings of rotating machinery. Adv. Mech. Eng. **8**(8), 1–21 (2016). https://doi.org/10.1177/1687814016664660
5. Wang, Q., Bu, S., He, Z.: Achieving predictive and proactive maintenance for high-speed railway power equipment with LSTM-RNN. IEEE Trans. Ind. Inf. **16**(10), 6509–6517 (2020). https://doi.org/10.1109/TII.2020.2966033
6. Wu, H., Huang, A., Sutherland, J.W.: Avoiding environmental consequences of equipment failure via an LSTM-based model for predictive maintenance. Procedia Manuf. **43**, 666–673 (2020). https://doi.org/10.1016/j.promfg.2020.02.131
7. Carvalho, T.P., Soares, F.A.A.M.N., Vita, R., Francisco, R.D.P., Basto, J.P., Alcalá, S.G.S.: A systematic literature review of machine learning methods applied to predictive maintenance. Comput. Ind. Eng. **137**, 106024 (2019). https://doi.org/10.1016/j.cie.2019.106024
8. Çınar, Z.M., Nuhu, A.A., Zeeshan, Q., Korhan, O., Asmael, M., Safaei, B.: Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. MDPI Open Access J. **12**(19), 1–42 (2020). https://doi.org/10.3390/su12198211
9. Jiménez, A., Zhang, L., Gómez Muñoz, C.Q., García Márquez, F.P.: Maintenance management based on machine learning and nonlinear features in wind turbines. Renew. Energy **146**, 316–328 (2020). https://doi.org/10.1016/j.renene.2019.06.135
10. Cho, S., May, G., Tourkogiorgis, I., Lazaro, O., de la Maza, B., Kiritsis, D., Perez, R.: A hybrid machine learning approach for predictive maintenance in smart factories of the future. In: IFIP AICT 536, APMS 2018, pp. 311–317 (2018)
11. Liao, L., Köttig, F.: A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. Appl. Soft Comput. **44**, 191–199 (2016). https://doi.org/10.1016/j.asoc.2016.03.013
12. Liu, Q., Liang, J., Ma, O.: A physics-based and data-driven hybrid modeling method for accurately simulating complex contact phenomenon. Multibody Syst. Dyn. **50**, 97–117 (2020). https://doi.org/10.1007/s11044-020-09746-w
13. Cao, Q., Samet, A., Zanni-Merk, C., de Bertrand de Beuvron, F., Reich, C.: Combining chronicle mining and semantics for predictive maintenance in manufacturing processes. Semant. Web **11**(6), 927–948 (2020). https://doi.org/10.3233/SW-200406
14. Christou, I.T., Kefalakis, N., Zalonis, A., Soldatos, J., Bröchler, R.: End-to-end industrial IoT platform for actionable predictive maintenance. IFAC-PapersOnLine **53**(3), 173–178 (2020). https://doi.org/10.1016/j.ifacol.2020.11.028
15. Fleischer, J., Klee, B., Spohrer, A., Merz, S.: Leitfaden Sensorik für Industrie 4.0—Wege zur kostengünstigen Sensorsystemen. VDMA Forum Industrie 4.0, Frankfurt (2018)
16. Celesti, A., Fazio, M., Márquez, F.G., Glikson, A., Mauwa, H., Bagula, A., Celesti, F., Villari, M.: How to develop IoT cloud e-health systems based on FIWARE: a lesson learnt. J. Sens. Actuator Netw. **8**(1):7, 1–24 (2019). https://doi.org/10.3390/jsan8010007
17. OMA: Homepage. http://www.openmobilealliance.org/release/ngsi/v1_0-20120529-a/oma-ts-ngsi_context_management-v1_0-20120529-a.pdf. Accessed 10/06/2021
18. Mathworks: Homepage. https://de.mathworks.com/company/newsletters/articles/three-ways-to-estimate-remaining-useful-life-for-predictive-maintenance.html. Accessed 10/06/2021

IoT Framework and Requirement for Intelligent Industrial Pyrolysis Process to Recycle CFRP Composite Wastes: Application Study

Mehar Ullah, Sankar Karuppannan Gopalraj, Daniel Gutierrez-Rojas, Pedro Nardelli, and Timo Kärki

Abstract The cumulating carbon fiber-reinforced polymer (CFRP) composite wastes need to be disposed efficiently. So far, the most effective thermal-based recycling technique, namely pyrolysis, has grown exponentially towards industrial scaling in developed countries such as the UK and Germany. Typically, even the slightest mistakes can cause unfavorable results and delays in workflow within such a massive operating environment (e.g., >1 tonnes/day operating capacity). The existing semi-automated and, in some cases, fully automated plants should be continuously updated to resemble the varying classes and volume of the CFRP composite wastes. To overcome such research gaps and imprecise manual errors, Internet-of-Things (IoT) based framework is proposed. This paper studies the theoretical implementation of an IoT-based framework into the pyrolysis process to recycle CFRP composite waste to manage the process based on the principles of cyber-physical systems. The proposed framework consists of sensors and actuators that will be used to collect the data and communicate with a central management constructed as a platform that will articulate and manipulate data to satisfy the requirements of the recycling process, computationally modeled through logical relations between physical entities. In this case, the management unit can be either controllable or monitored remotely to increase the operation time of the plant. Our objective is to propose a scalable method to improve the recycling process, which will also help future decision-making in handling recycled carbon fiber. Specifically, this study will go beyond the state-of-the-art in the field by (i) automatically calculate the mass of the waste and adjust the operating time, temperature, atmospheric pressure, and inert gas flow (if needed), (ii) regenerating heat so that after the first batch is recycled, the resin high in calorific value will be burned and will be releasing energy, whose generated heat needs to be trapped inside the furnace and then regenerated into the system, and (iii) decrease energy consumption and fasten the process flow time. In summary, the proposed framework

M. Ullah (✉) · D. Gutierrez-Rojas · P. Nardelli
School of Energy Systems, LUT University, Yliopistonkatu 34, 53850 Lappeenranta, Finland
e-mail: Mehar.Ullah@lut.fi

S. K. Gopalraj · T. Kärki
Fiber Composite Laboratory, Department of Mechanical Engineering, LUT University,
P.O. Box 20, 53850 Lappeenranta, Finland

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_26

aims to provide a user-friendly control and temperature monitoring that can increase the overall efficiency of the process and avoid possible process shut down or even char formation by a controlled atmosphere in the pyrolytic reactor.

Keywords IoT framework · Industrial-scale pyrolysis · Recycling carbon fiber

1 Introduction

Carbon fiber-reinforced polymer (CFRP) composites have been exponentially used in high-performance applications for decades. The composites have high mechanical properties for a lower weight ratio making them capable of replacing traditional metals in lightweight applications. However, CFRP composites employed 20 years ago have now reached their end-of-life (EoL) and raised a significant question about their disposal routes. So far, 62,000 tonnes of CFRP composite wastes have been cumulating each year, and the forecast predicts that the amount could increase up to 90,000+ tonnes/year if not disposed properly. At the same time, the annual demand for virgin CFRP composites also expected to be increased from 72,000 to 140,000 tonnes/year [1]. To establish a balance, recycling the waste composites, recovering the valuable carbon fibers (CFs), and reusing them into new composites is the only sustainable option [2].

Previously, landfilling and incineration were the popular disposal methods for their composites. However, various studies have proved that recycled carbon fibers (rCFs) to be close to their virgin properties, recycling industries have invested in a sustainable alternative to recycling CFRP composite wastes. Thermal recycling processes such as pyrolysis and solvolysis using supercritical/subcritical water or mild solvents have proved to be highly efficient [1]. Recently, a novel thermal recycling process [2] capable of recycling CFRP composite wastes with maximum efficiency has resulted in clean recycled fibers without disturbing the fibers' structural integrity (fiber direction, arrangement, and length). However, all these processes exist on a laboratory scale. Among them, pyrolysis has been successfully established on an industrial scale in The UK (Gen 2 Carbon), Belgium (Procotex), and Germany (CFK valley and SGL group).

Figure 1 presents the operating principle of the pyrolysis recycling process to recover CFRP composite wastes. First, the composite waste is size reduced using mechanical shredding and feeds into the system. The pyrolysis reactor is a closed chamber with no oxygen present. However, the process is done in an inert gas atmosphere to separate the valuable CFs from the matrix at 550 °C for the required time (depends on the waste quantity). Then, the recovered fibers are passed to a secondary heating chamber at 200 °C to oxidize the resin residues. Finally, the CFs are recycled, acquiring pyrolytic oil (can be used as feedstock) and hot gas (can be regenerated) [1]. Overall, industrial-scale pyrolysis possesses enormous sustainable benefits.

The EU's regulations have moved from landfilling and incineration towards sustainable recycling regulations. In which recycling processes with low emissions

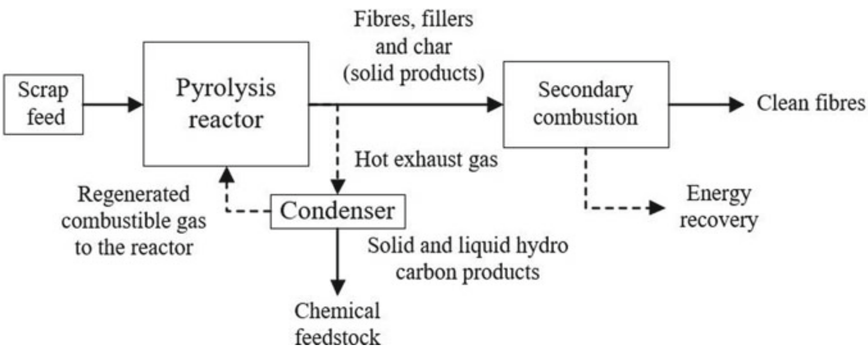


Fig. 1 Overview of the pyrolysis process [1]

in carbon footprints are preferred [1]. So, industries utilizing the pyrolysis process to recycle CFRP composite waste have to reconsider advanced emission monitoring along with their primary monitorable parameters such as pyrolysis temperature, pressure, inert gas flow, heat regeneration, recycled fiber quality, etc. Also, considering the type of CFRP composite wastes cumulating, which lacks composite information such as profile, the volume of fiber and resin, composite type, resin type, etc., there is a need to implement an advanced monitoring system to moderate such industrial-scale process.

These strategies call for more control and automation. Therefore, applying IoT scheme frameworks can improve the overall connectivity bringing significant control, monitoring, and safety to the process [3]. However, the existing semi-automated and, in some cases, fully automated pyrolysis plants need to be continuously updated and constantly monitored to couple with the varying CFRP composite waste types. To overcome such research gaps and eradicate manual errors, Internet-of-Things (IoT) based framework is proposed. Typically, an IoT is a network of devices interconnected to each other using some communication technology and using sensors and actuators to gather data from different devices and send that data to the cloud to store, process, and get information [4].

This paper studies the theoretical implementation of an Internet-of-Things (IoT)-based framework into the pyrolysis process to recycle CFRP composite wastes and manage the process based on the principles of cyber-physical systems. The proposed framework consists of sensors and actuators to collect the required data and communicate with central management constructed as an IoT platform that will articulate and manipulate data to satisfy the requirements of the recycling process, computationally modeled through logical relations between physical entities. Furthermore, this study will focus on selecting a suitable IoT platform based on the requirements of the pyrolysis process. Incorporating IoT and its platform into the pyrolysis process will improvize the plant by automatically calculating the mass of the CFRP composite wastes and adjusting parameters such as: operating time, temperature, atmospheric pressure, inert gas flow, increase energy efficiency, and reduce process flow time.

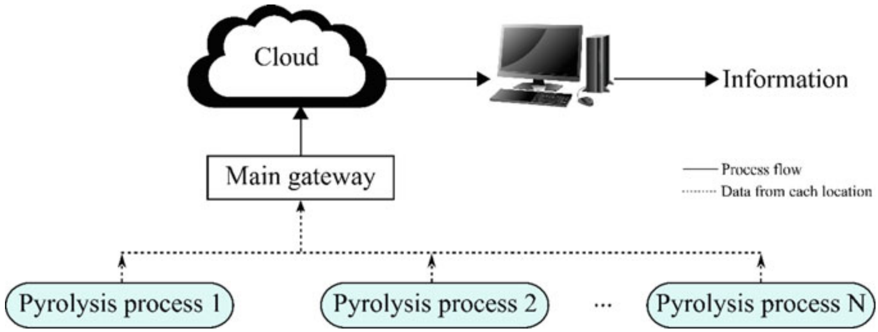


Fig. 2 IoT framework implementation in industrial-scale pyrolysis

2 General IoT Framework for Recycling Carbon Fibers

The focus on implementing IoT technology into industrial-scale pyrolysis was discussed in the following section. First, an insight into improving pyrolysis factors such as cost, speed, reliability, scalability, and performance resulting from IoT was investigated [5]. Furthermore, discussions on IoT collected data as a reliable source of information in current decision-making and future improvements into the pyrolysis process.

2.1 *Industrial-Scale Pyrolysis Process and IoT Framework*

Figure 2 presents the implemented IoT network in an industrial-scale pyrolysis setup. The process includes sensors and actuators at all the crucial sections (starting from the waste feed till the CF recovering) throughout the process. The IoT framework is designed to be flexible and extended to multiple furnaces (pyrolysis process 1, pyrolysis process 2 … pyrolysis process N). Figure 1 presents the pyrolysis process setup. The gathered data from the individual furnaces are sent to the main gateway and forwarded to the cloud for storage, processing, and visualization. In addition, information can be taken from data stored in the cloud. Based on that data, a further future business decision can be taken that can improve and speed up the CF recycling process with precision and accuracy.

2.2 *The Role of IoT in Process Monitoring*

CFRP waste recycling process will be benefited from an IoT deployment by adding physical elements to the cyberworld on what is known as a Cyber-Physical System (CPS), as introduced in [4]. This will enable better control of the input material

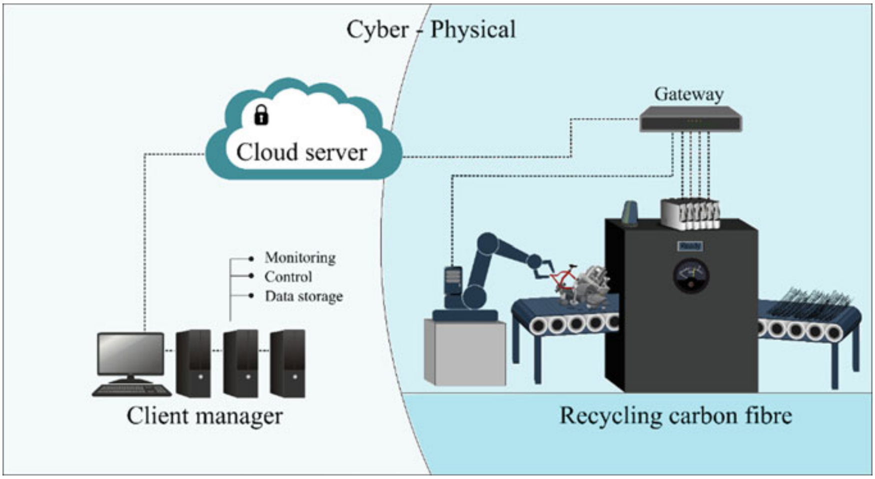


Fig. 3 IoT schematic of an in-site CF recycling process

transported along with the conveyor, control temperature of the furnace, and real-time monitoring of char production. In Fig. 3, it can be seen how IoT is used to form a cyber-physical environment where sensing and actuators are connected to the gateway making the interface to the cloud server. In the cyber world, a manager can remotely perform monitoring, control, or analyze stored data.

2.3 Selection of IoT Platform for Pyrolysis Process

Figure 4 presents the framework of IoT platforms. IoT application needs a platform to run smoothly and provide the data to make future decisions based on the data received from the IoT platform. Hundreds of IoT platforms are available, and finding a suitable IoT platform for a specific IoT application is complex. A lack of experience and knowledge compounds the problem, and in some cases, a company may select a platform without adequate requirements analysis, which later leads to problems [6]. Companies can select an appropriate IoT platform for their IoT application if they first analyze their business requirements and start selecting the IoT platform with precise business requirements and knowing key factors of IoT platforms [5].

The process for the selection of an IoT platform has five stages. In stage 1, the company identify and finalizes their business requirements. In stage 2, the identified requirements are categorized as “Important” or “Required”. In stage 3, the requirements are compared with the features provided by the IoT platforms. Only the five most important IoT platforms based on the market shares are selected in this case. In stage 4, IoT platforms are selected that are capable of fulfilling the requirements. Finally, in stage 5, a suitable platform is selected for the business IoT application.

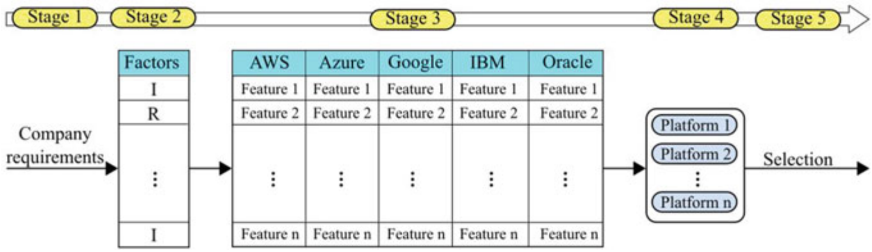


Fig. 4 Framework for the selection of the IoT platform

The requirements for the pyrolysis process to recycle CFRP composite waste are to eliminate the manual work involved in the process, speed up the recycling time (>1 tonnes/day), increase energy efficiency, and increase the overall recycling efficiency of the process. Furthermore, it calculates the amount of heat required for a specific amount of CFRP composite waste and estimates the amount of heat generated during the process. Finally, calculates the amount of emissions (exhaust and later outside the system). In short, the pyrolysis process needs stability, flexibility, scalability, security, attractive interface, data analytics, and interoperability throughout the system process. The process also needs and user-friendly application development environment for its IoT business application in the cloud.

The requirements of the pyrolysis process are categorized into "Important" and "Required" factors. The requirements stability, flexibility, scalability, security, attractive interface, and data analytics are considered "Important", and the requirements interoperability is considered "Required". The requirements of the pyrolysis process are compared with the features provided by the selected five major IoT platforms, as shown in Table 1. All the essential requirements of the pyrolysis process were matching with the features provided by the AWS IoT platform. However, Azure lacked the required flexibility and the required factor Interoperability.

2.4 Communication Requirements

Currently, wireless communication advances have countless open opportunities for industrial applications. They can be critical enablers for monitoring and controlling impossible tasks before due to low flexibility and cost [7]. For the past decades, requirements for the industrial-scale process have been discussed, but they are bounded to every application. 5G cellular communication, evolution from previous 4G networks, has come with an improved set of characteristics that can increase the operational performance of industrial applications. Despite all the advantages, challenges such as interoperability, quality of service, ease of use, reach, cost, and security remain essential goals to be investigated to ensure overall benefit [8].

Table 1 IoT platform features adopted from [5]

Factors	AWS	Azure	Google Cloud	IBM Watson	Oracle IoT
Scalability	Yes	Yes	Yes	Yes	Yes
Flexibility	Yes	–	Yes	–	Yes
Data analytics	Yes	Yes	Yes	Yes	Yes
Disaster recovery	Yes	Yes	No	No	No
Stability	Yes	Yes	Yes	–	–
Security	High	High	High	High	High
Data ownership	–	Yes	–	–	–
Protocol support	Yes	Yes	–	Yes	Yes
System performance	Yes	–	Yes	Yes	–
Time to market	Yes	Yes	–	–	Yes
Legacy architecture	Yes	–	–	–	Yes
Attractive interface	Yes	Yes	–	No	–
Pricing model	Bad	Bad	Good	–	–
Cloud ownership	Yes	Yes	Yes	–	Yes
Interoperability	Yes	–	–	–	Yes
App. environment	Yes	Yes	Yes	Yes	Yes
Hybrid cloud	Yes	Yes	–	–	–
Platform migration	Yes	Yes	–	–	–
Previous experience	Yes	Yes	–	–	–
Edge intelligence	Yes	Yes	Yes	–	Yes
Bandwidth	–	–	Good	–	–

5G technology achieves superiority due to its main three cornerstones: massive machine-type communications (mMTC), ultra-reliable and low latency communications (URLLC) Enhanced mobile broadband (eMBB) [3]. In Fig. 2, the communication means can be achieved by 5G wireless communication in the recycling process environment. The main argument for choosing this is scalability, flexibility, and cost. A wired communication system will need physical installation architecture to increase the cost of a component's change significantly. The communication requirements for the recycling process are as follow: (a) update time (process data) 1.5 ms, (b) transmission time (process data) 0.5 ms, (c) distance between logical endpoints 250 m, (d) reliability (redundancy), (e) less than 50 devices connected at the same time. The requirements can be achieved by installing one 5G base station inside the recycling building to ensure full connectivity between devices and cloud servers and allow remote operability.

3 Conclusion

The industrial-scale pyrolysis process to recycle CFRP composite wastes was studied by implementing IoT technology. The study proposes implementing sensors and actuators to collect data from the recycling plant and keeping in mind that the incoming composite wastes to be recycled will have varying composite types and properties. Therefore, the proposed IoT framework will adopt the recycling process conditions according to the composites to unify the standard pyrolysis setup. Furthermore, the methods involved in selecting a suitable IoT platform for the data collection, processing, storage, and visualization of results were also discussed. Overall, implementing the proposed IoT framework will enhance the recycling process adaptability over various CFRP composite waste types. As a result, the proposed IoT framework will facilitate the pyrolysis process into efficient energy utilization, reliability, scalability, and reduce the overall recycling time and cost. Furthermore, the information generated from the data collection and processing using the IoT platform can monitor and maintain the pyrolysis plant. Also, it is capable of influencing current and future recycling business decisions.

References

1. Karuppannan Gopalraj, S., Kärki, T.: A review on the recycling of waste carbon fibre/glass fibre-reinforced composites: fibre recovery, properties and life-cycle analysis. SN Appl. Sci. **2**, 433 (2020). https://doi.org/10.1007/s42452-020-2195-4
2. Gopalraj, S.K., Kärki, T.: A study to investigate the mechanical properties of recycled carbon fibre/glass fibre-reinforced epoxy composites using a novel thermal recycling process. Processes 2020 **8**(8), 954 (2020). https://doi.org/10.3390/PR8080954
3. Narayanan, A., Sena, A.S.D., Gutierrez-Rojas, D., Melgarejo, D.C., Hussain, H.M., Ullah, M., Bayhan, S., Nardelli, P.H.J.: Key advances in pervasive edge computing for industrial internet of things in 5G and beyond. IEEE Access **8**, 206734–206754 (2020). https://doi.org/10.1109/ACCESS.2020.3037717
4. Gutierrez-Rojas, D., Ullah, M., Christou, I.T., Almeida, G., Nardelli, P.H.J., Carrillo, D., Sant'Ana, J.M., Alves, H., Dzaferagic, M., Chiumento, A., Kalalas, C.: Three-layer approach to detect anomalies in industrial environments based on machine learning (2020)
5. Ullah, M., Nardelli, P.H.J., Wolff, A., Smolander, K.: Twenty-one key factors to choose an IoT platform: theoretical framework and its applications. IEEE Internet Things J. **7**, 10111–10119 (2020). https://doi.org/10.1109/JIOT.2020.3000056
6. Ullah, M., Smolander, K.: Highlighting the key factors of an IoT platform. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 901–906. IEEE (2019)
7. Underberg, L., Willmann, S.: Categorization of industrial communication requirements as key to developing application profiles. IFAC-PapersOnLine **53**, 8297–8302 (2020). https://doi.org/10.1016/J.IFACOL.2020.12.1919
8. El Khaled, Z., Mcheick, H.: Challenges and requirements for communication technologies in connected zones. Procedia Comput. Sci. **155**, 27–34 (2019). https://doi.org/10.1016/J.PROCS.2019.08.008

Approach of Automated ML Algorithm Selection for the Realization of Intelligent Production

Johannes Wimmer, Carmen Constantinescu, and Bastian Pokorni

Abstract For achieving the ambitious objectives of intelligent production, the artificial intelligence through its machine learning algorithms represents one of the most promising technology. The employment of machine learning algorithms for the optimization of complex production processes faces the big challenge of selecting the suitable machine learning method which fits the optimization parameter objectives. This paper introduces our approach for automated selection of ML algorithms to be used for optimization of a specific production process. The approach and the corresponding method consists of the following main blocks: (a) production process definition, (b) ML performance, (c) selector constructor and (d) assessment and incremental improvement of selector performance. The first component defines a typical production process or domain based on a well-established set of features, e.g. product quality inspection through process features as accuracy, material characteristics, etc. The ML performance construct contains precise defined performance of well clustered ML algorithms based on established benchmarking. The third construct, the Selector, automatically realizes a perfect mapping between the production process features and the performance of the ML algorithms. The logic of this automated selection represents the innovation of our work. The last component assesses the selector algorithm based on a set of specific KPIs for each production domain or process. The incremental improvement of the selector is approached as well, closing the loop between all method components. The developed approach and method have as foundations our work on identifying critical production processes/domains as core of realizing the intelligent production and laborious developed collection of ML algorithms, based on their performance data. These foundations and a motivation scenario are presented inside our paper to highlight the relevance of our research work.

Keywords Intelligent production · Artificial intelligence · Machine learning · Algorithm selection · AutoML

J. Wimmer (✉)
University of Stuttgart IAT, Stuttgart, Germany
e-mail: johannes.wimmer@iat.uni-stuttgart.de

C. Constantinescu · B. Pokorni
Fraunhofer IAO Stuttgart, Stuttgart, Germany

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_27

1 Motivation, Research Gap and Problem Statement Towards Intelligent Production

Artificial Intelligence (AI) in production is finding more and more widespread application. Be it in maintenance, logistics, quality control or production control and planning. The use cases are diverse and promise great potential for companies that opt for AI in their production. The basis for this is available data and sufficient basic process and AI knowledge. Based on Big Data and the right experts, a wide range of applications can be implemented today. However, companies still find it difficult to identify the right use cases and to develop initial prototypes independently. They often lack the expert knowledge and a general approach to identify and implement AI use cases. Initial procedures like AutoML already exist but are still at a very high level and do not lead the user step by step to the final application [1]. There is a need for action especially in the area of translating industrial needs to AI use cases and find the right algorithm and models for it without data scientist and AI experts [2]. Solutions are either provided by ready-made algorithms and frameworks or, in science, one helps oneself with try and error and tries to find the right algorithm for the specific use case through targeted trial and error and comparing different models and their results. In practice, however, this costs a lot of time and requires expert knowledge that companies often do not yet possess. Both of them lead to high cost for AI cases in production these days. In the course of intelligent production, it will nevertheless be inevitable that production engineers will be able to identify use cases independently and also implement them cost- and timesaving. For this purpose, a methodology is to be developed that specifically helps in the selection of the appropriate machine learning algorithm and thus specifically supports the companies in implementing and piloting their own AI use cases.

Research into intelligent production and the targeted use of AI differs from existing research into AI in that it is not about research into new fields of application for AI in production, as is already being done in many cases today, but about targeted research into the introduction of AI in production. This new approach requires a targeted focus on selecting the right algorithm for the diverse solutions in the individual areas of smart production [3]. Wherever data is generated or can be generated in production, a production engineer should be able to introduce an AI application in the future. AutoML procedures are already intended to provide assistance in this regard for non-experts in the field of AI [4]. However, these procedures do not cover the identification of new AI use cases and thus can only be used partially. The new approach to be researched includes and combines both the identification of new AI use cases and the selection of the most important features of the application and the appropriate algorithm. This should lead to a higher use of AI in production and faster development of new cases for the intelligent production of the future.

The remainder of the paper is organized as follows. Related Work and State-of-the-Art is summarized in Sect. 2. Section 3 presents the new approach and framework on AI Algorithm Selectors and Use Case identification in the intelligent production of the future. Section 4 concludes the paper with a short summary and gives an outlook on future work and planned next steps.

2 State-of-the-Art and Related Work

In this section the state-of-the-art in the domain of AI and Machine Learning and valuable approaches which we consider relevant for our work will be described.

2.1 Artificial Intelligence

Since the field of **AI** is complex and influenced by different disciplines, a clear definition of AI fields proves to be difficult. There is no uniform definition among AI researchers, and there is as yet no agreement on the most efficient approach to developing artificial intelligence [5]. Some experts focus on the IT aspect. This gives rise to AI domains such as “modelling”, “problem solving”. However, such classifications are not very useful, as they tend to focus on the fundamentals of AI rather than the areas of use. The four main research areas of AI are currently represented are language processing, image processing, expert systems, and robotics. The boundaries between the areas of AI are disappearing more and more. For example, all four areas are relevant for the use of an autonomous vehicle [6].

Next, we want to describe shortly the main sub domains of AI and current trends in this field.

2.1.1 Machine Learning and AutoML Procedure

Machine learning is an essential area of AI and is so popular that the term is often used as a synonym for AI. Basically, it is about improving learning based on data. The core of the field is testing the extent to which tasks are continuously solved better by the machine based on good training data or particularly large amounts of data from algorithms [7]. In the field of machine learning, algorithms are used which are capable of learning independently and optimizing themselves accordingly. An algorithm is a programmed command that processes the input data in a predefined form and leads to the output of results based on it. In machine learning, so-called self-adaptive algorithms are used so that the machines can learn independently without having to intervene in the ongoing learning process. Large amounts of data are necessary to train the algorithms in such a way that they are able to master predefined tasks better and better without having to be programmed repeatedly. In order to promote these learning processes, extensive, high-quality data sets are necessary as training material. These training data are used to generate the new algorithms. After completion, these are continuously checked with further input data to arrive at better decision-making bases. In machine learning, different learning styles are distinguished from each other [2].

Until now, highly qualified, and experienced experts in the fields of ML algorithms and statistics were necessary to develop an ML application [8]. With AutoML, feature development and further steps of model training are performed automatically and with a high degree of accuracy [9]. In addition to static grid and random search methods, much more efficient evolutionary and Bayesian methods are also used [10]. Thus, ML applications also become directly accessible to domain experts without being dependent on a Data Scientist [8, 11].

2.1.2 Supervised and Unsupervised Learning

In supervised learning, both the data set and the correct answer choices are already known. The algorithm only needs to be adapted so that the correct relationship between input and output data can be established. In the field of supervised learning, algorithms of classification as well as regression analysis are used [5].

In unsupervised learning, the system has no predefined target values and is supposed to identify patterns in the data on its own. The user has no prior knowledge about the similarity of the data. It is the algorithm's job to identify patterns on its own. Therefore, the insights gained by the AI system may be beyond human comprehension. In the context of unsupervised learning, the algorithm is given unlabeled data and is expected to recognize a structure on its own. To do this, the algorithm determines groupings of data that have similar behaviour or properties. An example of such a use case is to identify people in social media (pattern recognition) who are susceptible to believing false information, commenting positively on it, and forwarding it [6].

2.1.3 Reinforcement Learning

In reinforcement learning, there is no ideal solution at the beginning of the training phase. The system has to determine solution paths independently step by step through a trial-and-error process in order to subsequently cancel or optimize them. This step-by-step process is accelerated by "rewards" (for solution ideas that work) as well as "punishments" (for approaches that do not work). This learning style is most often used when there is limited training data or when the ideal outcome is not clearly definable. In addition, one uses procedures of reinforcing learning when something can only be learned from the interaction with the environment [5]. Reinforcement learning differs from supervised learning in that in supervised learning a training data set with labeled examples is provided by an external person and learned from. The goal of this type of learning is for the system to extrapolate its outputs to function correctly in situations that are not included in the training set. This behaviour is not sufficient for learning from interactions. In interactive problems, it is often impractical to obtain examples of the desired behaviour that are both correct and representative of all situations in which the system must act. In unfamiliar territory, a system must learn from its own experience. One of the current challenges of this discipline is the trade-off between investigation and exploitation. An example of augmented learning is Google's AlphaGo computer program. This program improved its strategy in simulated games of the board game against itself so that it ended up defeating one of the strongest players Lee Sedol [5].

2.1.4 Deep Learning

Deep learning is a subset of machine learning. The word "deep" refers to the number of layers in a neural network. Special networks are developed for this purpose that can take in large amounts of input data and process it within several layers. In contrast to simple neural networks, deep neural networks use special optimization methods that have a more extensive internal structure. Deep features and correlations are identified that relate the existing data points to each other. In Deep Learning, larger amounts of data can be processed and often more accurate results can be obtained than with traditional machine learning approaches [6]. In both Deep and Reinforcement Learning, the AI system learns autonomously. The difference between the two types of learning is that in deep learning, the AI system learns from a training set and then applies this learning to a new data set, while in reinforcement learning, the AI system learns dynamically by adjusting actions based on continuous feedback in order to maximize a reward [12].

2.2 Data Mining Models and Employment of Artificial Intelligence for Achieving Production Excellence

The CRISP-DM (Cross-Industry Standard Process for Data Mining) model was developed in the mid-1990s by industry leaders and endorsed by more than 200 data mining users and service providers. The model is free, documented and non-proprietary. With the goal of promoting best practices in organizations, this model provides the industry with the necessary framework to deliver projects with better quality and speed. The CRISP-DM model breaks down the data mining process into 6 phases: Business Understanding, Data Understanding, Data Preparation, Model Building, Testing and Evaluation, and Deployment [13].

Another methodology that is also used in a data mining study is SEMMA, which was developed by SAS Institute and divides the process into 5 phases, "Sample, Explore, Modify, Model and Assess" [14]. By following and correctly performing the phases of either of the above methods, members of the AI project can better understand the difficulties of the project and its planning. It also creates a roadmap so that it is easier to monitor the progress of the project.

In Germany, only 10% of manufacturing companies are using AI, according to a 2019 study by Fraunhofer IAO. Another 11% were in the process of preparing the introduction of AI in their company. However, just under 80% of companies had not yet launched their own initiative in this area or were currently in the process of assessing the potential [15]. The main use cases in Germany are in the area of intelligent assistance systems (413 use cases) and predictive analytics (369), as well as intelligent automation (320), according to a study by "Plattform Lernende Systeme" from 2021 [16].

Studies also show that small and medium-sized enterprises, as well as industry in general, often lack qualified specialists and the necessary AI expertise to identify and implement their own use cases [17]. This is precisely where the two approaches mentioned above are intended to support the experts, but they only very strongly describe the "what" and not the "how". So there is still a need for action here, particularly in the systematic selection of use cases and the selection of the right AI model, in order to find more AI applications in the intelligent production of the future [18].

3 Approach and Reference Model for Intelligent Production

The new approach to identify AI use cases in production and automatically selecting the appropriate algorithm towards Intelligent Production involves four steps as follows [19, 20]:

1. The selection of the use case and the production area as well as the associated processes and mechanisms of action. In addition, the objective of the use case should be defined here.
2. The selection of the appropriate ML method for the application area and concrete use case from an ML cluster and the corresponding mechanisms of action. In addition, it should be determined here whether and how much data is available for use or whether additional data must be generated.
3. The selection of the best and most performant algorithm in the selected ML class from the available set of algorithms for the given objective. This step represents the main work in the further research project.
4. The evaluation of the quality and performance of the selected and identified algorithm and the incremental improvement of the selection system, if necessary.

The above mentioned four steps together with its component elements are presented in Fig. 1, by highlighting in diverse colours their internal structure. The individual steps consist of following elements:

Step 1: Production Area Selection:

- Guiding questions for a clear description of the use case;
- Classification of the use case in a factory level of intelligent production;

Fig. 1 Steps of the new approach and their structure

- Classification in generic application scenarios;
- Definition of the concrete objective of the use case.

Step 2: ML-Method Selection:

- Question and requirement profile of the concrete use case;
- Mechanisms of action of the application and check if and how much data is available;
- Matching the objective of the application with the mechanisms of action of the existing ML methods;
- Final selection of the correct ML-Method.

Step 3: ML-Model and -Algorithm Selection:

- Clustering of possible algorithms according to data quantity and accuracy;
- Selection of the top and most performant algorithm or Best-Practice-Algorithm as Benchmark for further improvements;
- Implementation of the application and algorithm.

Step 4: Quality and Performance Evaluation:

– Measure the results of the implementation in terms of accuracy and error rate in the application;
– Incremental improvement of the use case and the algorithm used.

4 Conclusion and Future Work

This paper describes the current situation of AI use cases in production and the associated procedure for identifying and implementing new use cases. Also, it was pointed out that new tools like AutoML are targeting to support industrial engineers and non-experts in the field of AI, but yet there is a complete approach missing how to identify and implement new use cases in the intelligent production step by step. A new approach for describing and finding suitable AI use cases in production was described and an approach and framework for the general procedure were presented.

The new approach is intended to help translate production problems quickly and precisely into AI problem definitions and to help select the correct and best-performing algorithm. This should save time and money and also enable non-AI experts to implement their own AI use cases in production and thus implement the intelligent production of the future.

Future work is planned to be the description of the generic use cases in production, defining well separated use case classes and levels to identify possible use cases and questions quicker and more precise. Also, the conception of specific guiding questions for the specific use cases and the associated data-oriented problem description should be done in a next step which should help to describe the industrial use case and translate it into a concrete ML task. Finally, one of the major points is the development of the concrete algorithm Selector for the intelligent production of the future. All of these aspects will be analysed, tested and evaluated in a planned demonstrator based on a research platform we use in ARENA2036 in the field of a cognitive digital twin combined with a quality measurement process in production and should be presented as a case study in the next paper.

References

1. Krauß, J., Pacheco, B.M., Zang, H.M., Schmitt, R.H.: Automated machine learning for predictive quality in production. Procedia CIRP **93**, 443–448 (2020). https://doi.org/10.1016/j.procir.2020.04.039
2. Seidel, R., Mayr, A., Schäfer, F., Kißkalt, D., Franke, J.: Towards a smart electronics production using machine learning techniques. In: 42nd International Spring Seminar on Electronics Technology (ISSE), pp. 1–6 (2019). https://doi.org/10.1109/ISSE.2019.8810176
3. Yao, Q., Wang, M., Escalante, H.J., Guyon, I., Hu, Y.-Q., Li, Y.-F., et al.: Taking human out of learning applications: a survey on automated machine learning (2018). Online verfügbar unter https://arxiv.org/pdf/1810.13306

4. Chauhan, K., et al.: Automated machine learning: the new wave of machine learning. In: 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), pp. 205–212 (2020). https://doi.org/10.1109/ICIMIA48430.2020.9074859
5. Gentsch, P.: Artificial Intelligence for Sales, Marketing and Service (2018)
6. Kreutzer, R.T., Sirrenberg, M.: Understanding Artificial Intelligence: Fundamentals—Use Cases—Enterprise AI Journey. Wiesbaden (2019)
7. Weber, F.: Artificial Intelligence for Business Analytics. Algorithms, Platforms and Application Scenarios, 1st edn. Springer Fachmedien Wiesbaden, Wiesbaden (2020)
8. Zöller, M.-A., Huber, M.F.: Benchmark and survey of automated machine learning frameworks. J. Artif. Intell. Res. **70**, 409–474 (2021)
9. Hutter, F., Kotthoff, L., Vanschoren, J.: Automated Machine Learning—Methods, Systems, Challenges. Springer, Cham, CH (2019)
10. Sakhnyuk, P.A., Sakhnyuk, T.: Intellectual technologies in digital transformation. IOP Conf. Ser. Mater. Sci. Eng. (2020)
11. Haricha, K., Khiat, A., Issaoui, Y., Bahnasse, A., Hassan, O.: Towards smart manufucturing: implementation and benefits. Procedia Comput. Sci. **177**, 639–644 (2020). https://doi.org/10.1016/j.procs.2020.10.091
12. Marr, B.: Artificial intelligence: what's the difference between deep learning and reinforcement learning? (2018). https://www.forbes.com/sites/bernardmarr/2018/10/22/artificial-intelligence-whats-thedifference-between-deep-learning-and-reinforcement-learning/#3ffdcadd271e
13. Kotu, V., Deshpande, B.: Predictive Analytics and Data Mining—Concepts and Practice with RapidMiner. Morgan Kaufmann, Waltham (2015)
14. SAS Institute: SAS® Enterprise Miner™ 14.3: Reference Help. 1. SAS Institute Inc., Auflage, Cary (2017)
15. Schuler, S., Hämmerle, M., Bauer, W.: Digitale Transformation—Gutes Arbeiten und Qualifizierung aktiv gestalten. In: Spath, D., Spanner-Ulmer, B. (Hg.) Digitale Transformation—gutes Arbeiten und Qualifizierung aktiv gestalten, S. 255–272. GITO (Schriftenreihe der Wissenschaftlichen Gesellschaft für Arbeits- und Betriebsorganisation), Berlin (2019)
16. Ecker, W., Coulon, C.-H., Kohler, M.: KI in die Anwendung bringen—Eine Gemeinschaftsaufgabe für Hochschulen, Forschungseinrichtungen Unternehmen und Politik. Whitepaper aus der Plattform Lernende Systeme, München (2021)
17. Deloitte: State of AI in the Enterprise, 3rd edn. Ergebnisse der Befragung von 200 AI-Experten zu künstlicher Intelligenz in deutschen Unternehmen (2020). https://www.deloitte-mail.de/custloads/141631293/md_1759242.pdf?sc_src=email_4411477&sc_lid=184874087&sc_uid=V2zf9Ae5Pe&sc_llid=427
18. Plattform Lernende Systeme: KI im Mittelstand—Potentiale erkennen, Voraussetzungen schaffen, Transformation meistern (2021). https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/PLS_Booklet_KMU.pdf
19. Applied AI: How to Find and Prioritize AI Use Cases. UnternehmerTUM GmbH, Garching (2020). https://aai.frb.io/assets/files/AppliedAI_Whitepaper_UseCase_Webansicht.pdf
20. VDMA Bayern: Leitfaden Künstliche Intelligenz—Potenziale und Umsetzungen im Mittelstand (2020). https://bayern.vdma.org/documents/106096/53103997/VDMA%2520Bayern_Leitfaden_KI_2020_1601889305004.pdf/a2460803-9ff9-9c00-15cf-aec87493705f

A Case Study on Data Analytics Based on Edge Computing for Smart Manufacturing System

S. K. H. Sit and C. K. M. Lee

Abstract Competitiveness of an electronics manufacturing services (EMS) firm is being closely monitored and interrelated in production cost control and product reliability performance level. Data analytics in massive manufacturing data can extricate huge business opportunities and values to the firms. Major challenge in data analytics application is heterogeneous and enormous of which dynamic data generated from continuously running production line reflects real-time velocity of production environment, so that the factory management demands data analytics to provide real-time solution for the improvement right on the spot. Cloud-based data analytics exhibits problems such as data capturing, storage, transfer latency and data quality that hinders the advancement of big data analytics in smart manufacturing horizon. Selection of appropriate data mining algorithm or techniques has been challenging to industry leaders in deriving desired patterns or model solving the exact problem they are facing. The aim of this research study is to illustrate the edge-based intelligent integrated information framework (INFO-I2) for the improvement of data quality in relevancy and enabling cloud-based data analytics to focus on product performance augmentation (Pipino et al. in Commun. ACM 45:211–218, 2002 [1]). In the case study, edge devices had been used for not only real-time data collection but also localized failure analysis and predictive analytics to perform autonomous decision-making in different workstations through production process. Cloud-based computing performs efficient optimization analytics for product functionality performance with those processed data which is an integrated production management system and knowledge base generated from localized data analytics of edge devices. The implementation of edge-based information framework improves the workflow

S. K. H. Sit · C. K. M. Lee (✉)
Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
e-mail: ckm.lee@polyu.edu.hk

S. K. H. Sit
e-mail: sheron.sit@connect.polyu.hk

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_28

management, eventually reduces manufacturing cost and improved product reliability. The contribution of this paper is to demonstrate how the proposed cloud-based manufacturing system architecture adopted both Cloud and Edge Computing to enhance product reliability and pave the way for smart manufacturing.

Keywords Smart manufacturing · Data analytics · Cloud-based · Edge-based computing

1 Introduction

Smart manufacturing creates competitive advantage in manufacturing industry. In a smart manufacturing environment, more and more production systems and devices are connected seamlessly to the Internet so that a large volume of data can be collected throughout all phases of production processes. In electronics manufacturing services (EMS) industry, most are enthusiastic about the manufacturing knowledge not only in product design such as the adoption of programme/software, product industrial design (ID), but also in production technology know-how as the whole capability package [2, 3]. In such circumstances, customers tend to intensively get involved in every stage of production process; gather as many data from manufacturing layers as possible; accumulate all production data back to their own centralized cloud storage, then perform the data analytics with data storage in the corporate data centre or centralized cloud services [3, 4].

Fundamental factors of success to an EMS in the daily operation are strategic cost management and quality yield management [5, 6]. Big Data analytics in smart manufacturing environment serves to better cost and quality control management, while major challenges to implement cloud-based integrated smart production management is the data processing ability of humongous expansion of data requiring the establishment and maintenance of a reliable cloud platform [7]. Quality of the data collected through production processes, connectivity of network from manufacturing site to data processing centre and instantaneity of the communication between manufacturing site and decision hub are other major challenges for those relying heavily on off-site centralized data analytics manufacturing system [1, 8].

Ren et al. [8] conducted comprehensive overview of big data in smart manufacturing and proposed a conceptual framework from the perspective of product lifecycle. To facilitate the new variety of applications and services to analyze a large volume of data and enable large-scale manufacturing collaboration, Qi and Tao [9] introduced a hierarchy reference architecture based on cloud computing, fog computing, and edge computing. Lee et al. [7] proposed an Industrial Internet of Things (IIoT) Suite, which consist of a Micro-services-based IIoT Cloud Platform and IIoT-based Smart Hub, to leverage the cost and emerging technology thresholds for industrialists in conducting industrial upgrading and transformation to achieve smart production and high value-added manufacturing processes.

Apart from smart manufacturing reference architecture, Hu et al. [10] focused on vertical integration of intelligent manufacturing by implementing an intelligent robot factory (iRobot-Factory), which adopted a highly interconnected and deeply integrated intelligent production line. Their experimental results showed significantly improvement in both assembly quality and the production efficiency, while the number of system instructions also decreased significantly. Though there are some open issues relating to cloud-end fusion, load balancing, and personalized robots to refer to promoting the emotion recognition and interaction experience of users, it's important to lay the foundations for future research with edge computing for smart manufacturing and case study helps practitioners and researchers to realize how the latest technology is adopted by industry.

2 Methodology

2.1 Big Data Analytics in Smart Manufacturing

The concept of smart manufacturing enables manufacturing companies to grasp their competitive advantage for making industry more efficient and sustainable. In modern manufacturing environments, more and more devices are connected to the Internet so that vast amounts of data are collected in database management systems and data warehouses during all phases of the product lifecycle such as product and process design, scheduling, materials planning, assembly, quality control, fault detection, equipment maintenance [11, 12].

It is widely recognized that conducting data analysis with massive manufacturing data not only can extract huge business values which enhance the competitiveness for the manufacturing companies, but also result in research challenges due to the heterogeneous data types, enormous volume, and real-time velocity of manufacturing data [13].

The application of big data analytics in the context of manufacturing processes and enterprises grew rapidly in recent years due to the wide adoption of Internet of Things (IoT). Choudhary et al. [14] stated that major data mining functions to be performed can be categorized in in five categories, characterization and description, association, classification, prediction, clustering and evolution analysis. Choosing the appropriate data mining algorithm, including the selection of techniques to perform the desired function to find the patterns in the data or deriving the model becomes challenges to industry leaders during the implementation of the latest technologies.

2.2 Cloud-Based Data Analytics and Edge Computing

The IIoT platform served as a base of application framework in dealing with the situation that huge quantities and numerous types of IoT devices in smart factories, warehouses and offices. However, the enormous extents of data exchange and communication, management, monitoring and control of IoT devices as well as the establishment and maintenance of a reliable cloud platform hinder industrialists to implement an integrated smart production management [15].

The IoT is a worldwide network of interrelated physical devices, which is an essential component of the internet, including sensors, actuators, smart apps, computers, mechanical machines, and people. Companies devoted to digital transformation take their future plants inspired by the IoT while computing capability becomes another limitation for them. Many research results on resource allocations in auction formats which have been implemented to consider the demand and real-time supply for smart development resources, but safety privacy and trust estimation issues related to these outcomes are not actively discussed. The existing resource assignment method in the smart production system cannot guarantee that resources meet the inherently complex and volatile requirements of the user are timely [16]. They demonstrated through experiments that when the IIoT equipment and gateways are effective, the utilization of each participant is improved to prove that effective allocation of the computing resources and the carrier is critical in the industrial internet of Things (IIoT) for smart production systems.

In the continuously running production system, the effective allocation of the computing resources and the carrier is critical in the industrial internet of Things (IIoT) for smart production systems [16]. The sooner the problem can be identified, the faster the mistakes can be corrected so that the less defectives can be produced. However, cloud based manufacturing information system faces fundamental problems, such as not enough bandwidth in the factory for large volume data transmission, latency created by different time-zone for customer and factory, network unavailability in the factory and delay related to security review on data transmission across boundary (e.g. headquarter in US and production facility in China), that restricts cloud-based computing information system in providing high-speed and low latency real-time applications to support continuously improvement in quality level of product reliability [15].

To achieve this, an integrated data analytics framework of using edge devices for real-time data collection, localized failure analysis, predictive analytics, and autonomous decision-making, and the cloud for aggregate data analysis, is proposed in this study (as shown in Fig. 1) [10]. Operation system optimization and trend analysis for overall product performance improvement was introduced in the case factory.

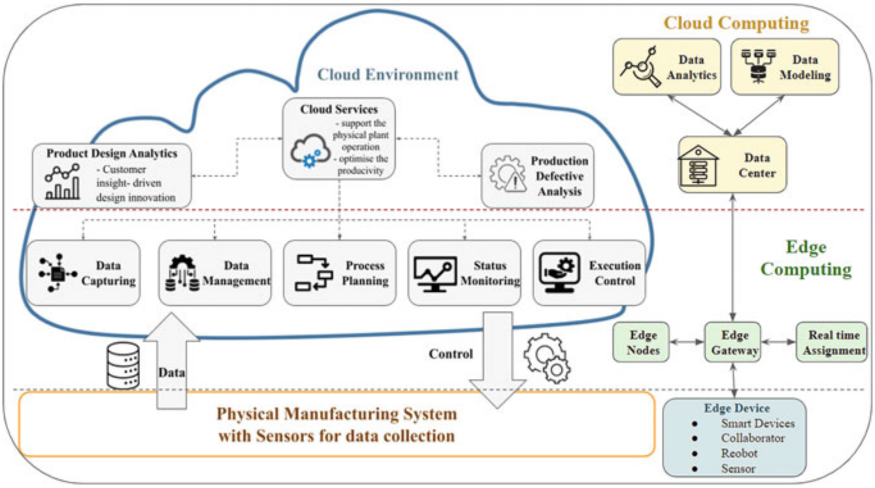


Fig. 1 The cloud-based manufacturing system architecture adopted both cloud and edge computing

3 Case Study

In Electronic Manufacturing Service (EMS) environment, the design of corporate level cloud-based data analytics aims to serve programme/software debugging in optimization of product design as well as production defective analysis to improve product reliability. Therefore, all data collected through physical manufacturing system are requested to be transferred and stored in the "data centre" of cloud computing manufacturing system shown in Fig. 1. As most resources of the cloud-based data analytic has normally been allocated to optimize and train the product operating system (software) to be smarter. Hence, analysing related to process or product reliability improvement have been put aside at lower priority in follow up. Those information normally take much longer time to arrive the factory comparing to software update instructions.

In the case factory, there are testing stations through the entire assembly process that generate massive data of defectives. The factory established the edge-based data analytics manufacturing system, and the new manufacturing system architecture is as shown in Fig. 2. Instead of sending all measurement data to be centralized, only processed data which is an integrated production management system and knowledge base including product reliability improvement and DFM solution, will be sent. Examination of these data concludes causes of defectives and facilitates the engineers to work out the improvement solutions consequently. In such situation, cloud-based computing performs efficient optimization for product functionality and performance evaluation.

There are three steps for the implementation of the edge-based data analytics manufacturing system, as demonstrated in Fig. 2.

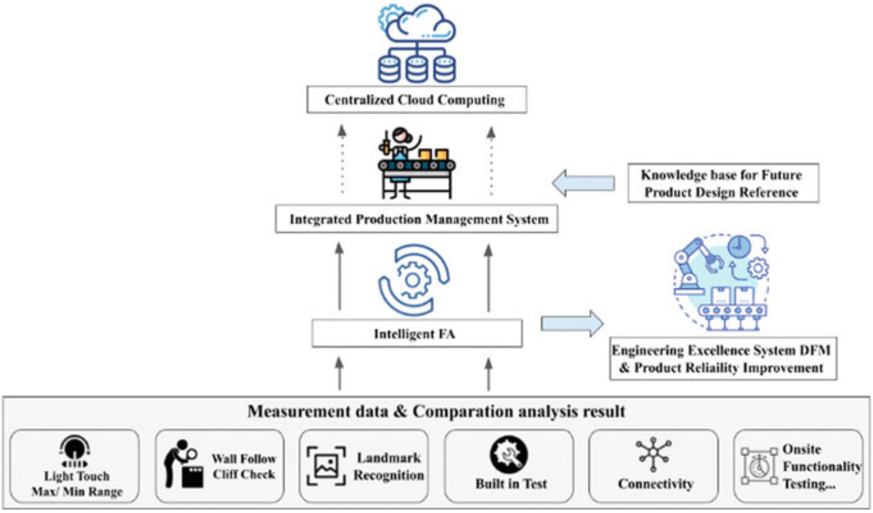


Fig. 2 The edge-based manufacturing system architecture (INFO-I2)

3.1 Digitalization of Data Collection in Testing Process

First, the factory converted manually operated testing stations into robotic/automated testing stations. Product performance measurements, such as computed calibration result performed by built-in programming in testing jigs had been collected and stored in the local server of testing robots. Real-time testing measurements and performance results (as shown in Fig. 3) in organized arrangement can be collected throughout testing processes.

```
computed-cliff-left-calibration min: 10 max: 150 computed: 10 PASS
computed-cliff-front-left-calibration min: 10 max: 150 computed: 61 PASS
computed-cliff-front-right-calibration min: 10 max: 150 computed: 55 PASS
computed-cliff-right-calibration min: 10 max: 150 computed: 23 PASS
computed-rear-cliff-left-calibration min: 10 max: 150 computed: 0 MARGINAL
computed-rear-cliff-right-calibration min: 10 max: 150 computed: 15 PASS
4719 5E5A 98A0 BFFF
LANGUAGE(S): english (0)
```

Fig. 3 Computerized testing measurement and performance result

	Computed-cliff											
	left:		right:		front-left:		front-right:		rear-left:		rear-right:	
1 4704 645D 98A0 CFFF	26	PASS	7	MARGINAL	59	PASS	57	PASS	0	MARGINAL	6	MARGINAL
2 4704 665E 98A0 CFFF	20	PASS	32	PASS	51	PASS	56	PASS	0	MARGINAL	0	MARGINAL
3 4705 5153 98A0 CFFF	9	MARGINAL	12	PASS	67	PASS	56	PASS	0	MARGINAL	10	PASS
4 4719 5D4B 98A0 BFFF	4	MARGINAL	66	PASS	53	PASS	58	PASS	0	MARGINAL	0	MARGINAL
5 4719 5E5A 98A0 BFFF	10	PASS	23	PASS	61	PASS	55	PASS	0	MARGINAL	15	PASS
6 4703 4D4D 98A0 CFFF	3	MARGINAL	34	PASS	39	PASS	47	PASS	0	MARGINAL	0	MARGINAL
7 4719 5B54 98A0 BFFF	17	PASS	32	PASS	52	PASS	56	PASS	5	MARGINAL	7	MARGINAL
8 4719 5C50 98A0 BFFF	25	PASS	32	PASS	58	PASS	63	PASS	11	PASS	0	MARGINAL
9 4719 5B53 98A0 BFFF	21	PASS	29	PASS	54	PASS	69	PASS	6	MARGINAL	8	MARGINAL
10 470B 644B 98A0 CFFF	15	PASS	17	PASS	53	PASS	53	PASS	1	MARGINAL	21	PASS

Fig. 4 Performance measurement for failure products

3.2 Localization of Data Analytics with Edge-Based Failure Analytics

The next step, massive testing data have been accumulated from production line, measurement figures on failure products had been extracted from the full list (as shown in Fig. 4).

Throughout concentrated failure analysis (FA) on individual defective units, a mapping showing correlation concerning failure and cause(s) has been summarized, mechanisms and causes of failures related to quality issues of product reliability found from field return had been identified. Analysis result shows that most failures are not extraordinary events, and most of those are foreseeable and preventable.

Edge-based computing focuses on intelligent FA to identify correlation between causes to failures through comparative data analytic based on massive performance measurement cumulated due to the use of automated testing fixture. Once any defectives been found in components, materials, design, and assembly had been found, the information will be inputted into the localized FA system in each testing fixture (edge node) as trained data to support future analytics. Through a train-and-learn process, information becomes knowledge base for the testing of the related production process.

3.3 Adoption of the Intelligent Integrated Information Framework (INFO-I2)

INFO-I2 framework includes data collection from individual testing fixture, localised data analytics, and data transmission to centralized cloud-based information system. This paper uses Integrated Mobility Testing (IMT) as example to describe architecture of the INFO-I2 framework and it enables data analytics for different purposes and efficiently improves product reliability in timely managed manner.

As shown in Fig. 5, Integrated Mobility Testing (IMT) is one of the 100% functional testing in production line. It integrates multiple testing process into one workstation. Multiple performance measurements have been recorded and stored in the local storage, known as edge nodes of this IMT testing fixture. The testing fixture then performs comparative analysis. According to predefined upper limit (UP) and lower limit (LL), the result of PASS or NG for every function been examined and alarms will be shown if there is a failure. By the time a result delivered to operator, measurements are stored in the testing fixture too.

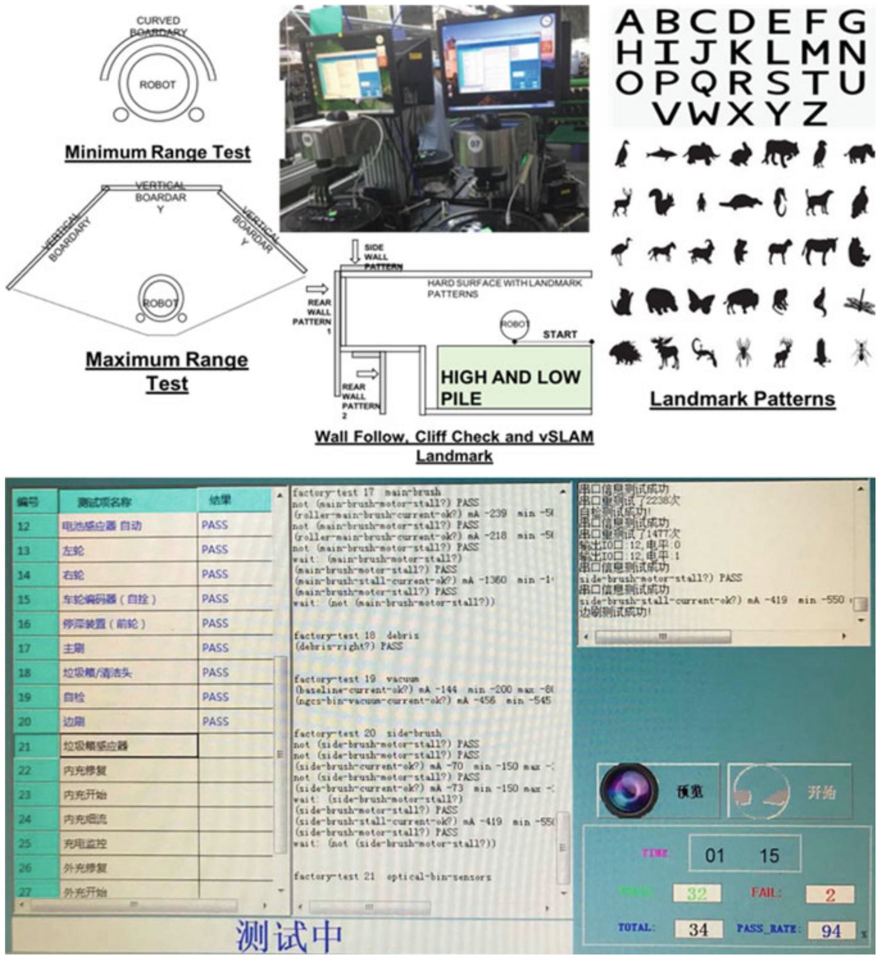


Fig. 5 Integrated mobility testing (IMT) architecture and measurements collected and stored in testing fixture storage

Knowledge bases in each node then has been integrated into a comprehensive production process monitoring guideline. The integrated knowledge base could be used in preventing further failure events and makes it possible to undertake corrective actions to optimize the product performance. In addition, the system guarantees production quality and reliability and optimize production cost as long run.

After the adoption of new manufacturing information system, the factory takes proactive actions in product reliability improvement; builds up engineering excellence knowledge base for future product and process design with analytics outcome of edge-based INFO-I2 framework from continuously mass production. This significantly improves product quality level and product reliability. Customer complaint drops out of top 3 of the ranking list within 6 months after the implementation.

4 Conclusion

To conclude with the motivation of the research, "Cost and Quality" is the leading criteria to a company to be sustainable in the manufacturing industry. "Production yield or Process yield" is important in particular at a multistage production chain such as EMS manufacturing system. Smart manufacturing and predictive analytics become vital to drive production yield to a new horizon. Big data analytics, cloud computing and edge computing in smart manufacturing system is important to drive the production into a vital smart manufacturing paradigm. Application of big data analytics through production and quality control process creates competitive advantage to manufacturing companies. The implementation of edge-based information framework improves the workflow management, eventually reduces manufacturing cost and improved product reliability. However, due to the resource limitation of edge severs, the conventional data analytics methods might be too complicated to be executed at edge servers. Therefore, the models of the data analytics methods need to be trained at remote cloud servers first and be transferred at local edge servers to enhance the processing capability. For future works, the authors suggest investigating lightweight data analytics schemes which can be deployed locally at edge servers approximate to users.

References

1. Pipino, L.L., Lee, Y.W., Wang, R.Y.: Data quality assessment. Commun. ACM **45**(4), 211–218 (2002). https://doi.org/10.1145/505248.506010
2. Fung, V.W.C., Yung, K.C.: An intelligent approach for improving printed circuit board assembly process performance in smart manufacturing. Int. J. Eng. Bus. Manag. **12**, 184797902094618 (2020). https://doi.org/10.1177/1847979020946189
3. Liu, C., Li, H., Tang, Y., Lin, D., Liu, J.: Next generation integrated smart manufacturing based on big data analytics, reinforced learning, and optimal routes planning methods. Int. J. Comput. Integr. Manuf. **32**(9), 820–831 (2019). https://doi.org/10.1080/0951192X.2019.1636412

4. Lee, J., Bagheri, B., Kao, H.-A.: A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manuf. Lett. **3**, 18–23 (2015). https://doi.org/10.1016/j.mfglet.2014.12.001
5. Mazzuto, G., Ciarapica, F.E.: A big data analytics approach to quality, reliability and risk management. Int. J. Qual. Reliab. Manag. **36**(1), 2–6 (2019). https://doi.org/10.1108/IJQRM-01-2019-294
6. Henri, J.-F., Boiral, O., Roy, M.-J.: Strategic cost management and performance: the case of environmental costs. Br. Account. Rev. **48**(2), 269–282 (2016). https://doi.org/10.1016/j.bar.2015.01.001
7. Lee, C.K.M., Zhang, S.Z., Ng, K.K.H.: Development of an industrial internet of things suite for smart factory towards re-industrialization. Adv. Manuf. **5**(4), 335–343 (2017). https://doi.org/10.1007/s40436-017-0197-2
8. Ren, S., Zhang, Y., Liu, Y., Sakao, T., Huisingh, D., Almeida, C.M.V.B.: A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: a framework, challenges and future research directions. J. Clean. Prod. **210**, 1343–1365 (2019). https://doi.org/10.1016/j.jclepro.2018.11.025
9. Qi, Q., Tao, F.: A smart manufacturing service system based on edge computing, fog computing and cloud computing. IEEE Access 1 (2019). https://doi.org/10.1109/ACCESS.2019.2923610
10. Hu, L., Miao, Y., Wu, G., Hassan, M.M., Humar, I.: iRobot-Factory: an intelligent robot factory based on cognitive manufacturing and edge computing. Future Gener. Comput. Syst. **90**, 569–577 (2019). https://doi.org/10.1016/j.future.2018.08.006
11. Qu, Y.J., Ming, X.G., Liu, Z.W., Zhang, X.Y., Hou, Z.T.: Smart manufacturing systems: state of the art and future trends. Int. J. Adv. Manuf. Technol. **103**(9–12), 3751–3768 (2019). https://doi.org/10.1007/s00170-019-03754-7
12. Shah, D., Wang, J., He, Q.P.: An internet-of-things enabled smart manufacturing testbed. IFAC-PapersOnLine **52**(1), 562–567 (2019). https://doi.org/10.1016/j.ifacol.2019.06.122
13. Dai, H.-N., Wang, H., Xu, G., Wan, J., Imran, M., Dai, H.N., Xu, G.: Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. Enterp. Inf. Syst. **14** (2019). https://doi.org/10.1080/17517575.2019.1633689
14. Choudhary, A.K., Harding, J.A., Tiwari, M.K.: Data mining in manufacturing: a review based on the kind of knowledge. J. Intell. Manuf. **20**(5), 501 (2008). https://doi.org/10.1007/s10845-008-0145-x
15. Varghese, B., Wang, N., Barbhuiya, S., Kilpatrick, P., Nikolopoulos, D.: Challenges and opportunities in edge computing (2016)
16. Xu, X., Han, M., Nagarajan, S.M., Anandhan, P.: Industrial internet of things for smart manufacturing applications using hierarchical trustful resource assignment. Comput. Commun. **160**, 423–430 (2020). https://doi.org/10.1016/j.comcom.2020.06.004

Supply Chain Management

An Integrated Approach for Resilient Value Creation Among the Lifecycle: Using the Automotive Industry as an Example

Florian Herrmann, Lukas Block, and Oliver Riedel

Abstract Current semiconductor shortages, trade obstructions, as well as the COVID-19 pandemic have shown the urgent requirement to rethink existing economic and production structures as well as trade practices. Such global events revealed the susceptibility of value-added structures to environmental disruptions. Especially the complex and highly optimized value chains in the automotive industry were particularly affected. Lack of early warning systems, missing redundancies and ever-increasing dependencies between individual value-added partners can be seen as some of the reasons for the enormous effects the industry is facing. However, looking into the future, fast and flexible product changes might affect the supply chain in similar ways as the current events. To secure value creation and employment and to minimize the impact of future crisis, the industries structures must be made more resilient. This paper provides a contribution to this through an integrated approach. This approach combines elements of linear value creation with a lifecycle-oriented perspective: The life cycle perspective considers temporal changes and dependencies between the individual supply chains of the life cycle phases. Subsequently, the value creation system is decomposed into manageable elements. The corresponding dependencies over selected life cycle phases are described. By creating scenario-robust value-added modules, risk within value-added systems is minimized. To illustrate the approach, a case study on automotive software and electronics is presented. Future changes regarding possible functions of an electric vehicle are anticipated and technical implications are derived on basis of a scenario analysis. Subsequently, the value-added module of integrated software and hardware development are examined in more detail and possible value-added configurations are displayed to indicate that the developed approach is suitable to generate robust supply chains. Interdependencies between value chains of multiple life cycle phases are discovered. Consequently, our approach provides additional value in the context

F. Herrmann (✉) · O. Riedel
Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering IAO, Stuttgart, Germany
e-mail: florian.herrmann@iao.fraunhofer.de

L. Block
Institute of Human Factors and Technology Management (IAT), University of Stuttgart, Stuttgart, Germany

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_29

of flexible products and services as well as their associated value-added processes and systems.

Keywords Automotive industry · Product life cycle · Value chain

1 Introduction

The topic of resilience has already been the subject of numerous studies in recent years ranging from resilience for urban climate systems, living organisms through materials to organizational leadership [1–3]. The topic has gained particular importance for corporate planning in the context of industrial research as well as in the context of optimizing production facilities and production networks. The concept of supply chain resilience acquired attention with the publications of Christopher and Peck [4] and Sheffi [5] in the early 2000s. Their work focuses on the risks to business in the context of wider supply chains and to approaches and strategies to reduce these risks. Building up on this they deal with the question how investments to achieve greater resilience can be transformed into competitive advantage. The fact that this concept became a permanent fixture in the area of supply chain management was driven by two factors according to Pettit et al. [3]. Firstly, business environment was getting more and more turbulent and volatile triggered by globalization and related corporate strategies Secondly, the traditional risk strategies of the companies were not sufficient to meet the high expectations in becoming a more resilient organisation. New analytical approaches and a constant shift in culture were needed [3]. Consequently, researchers all over the world started to address a wide range of related research questions to make organizations and value networks more resilient. The research work covered topics such as the categorization of the capabilities and vulnerability, their interdependencies or the analysis of factors that drive resilience improvement [6, 7]. Further topics are resilience in connection with product innovativeness and performance on a corporate level, the resilience of product systems or the role of collaborations to strengthen supply chain resilience [8–10]. The approaches and methods applied for these purposes include life cycle assessments and a broad diversity of modelling and simulation techniques (Discrete event simulations, agent-based simulations or Petri-net modelling) [3, 9]. In addition to the previously sketched research fields various definitions for the terminus resilience were produced [4, 11, 12]. A definition, which can serve as a generic approach in a business context was for example made by Fiksel [12] who understands resilience as “the capacity for an enterprise to survive, adapt, and grow in the face of turbulent change”. Others highlight within their definitions elements like resistance and recovery in relation to occurring disruptions or take an even stronger system perspective [3].

2 The Automotive Industry in the Wake of COVID-19

COVID-19, the aftermath of the "Ever Given" obstruction in March 2021 and the following semiconductor shortage revealed the susceptibility of value-added structures to environmental disruptions. This is especially true for the complex and highly optimized value chains in the automotive industry [13]: An analysis of a consulting company suggests, that the global automotive industry will produce about four million vehicles less in 2021 due to such disruptions [14].

However, not only the availability of components leads to challenges in the automotive software and electronic supply chain. Unknown security vulnerabilities in microcontrollers, unforeseen performance spikes (e.g. in communication gateways) and future function extensions might lead to a changes in automotive design. Ultimately, the design and thus the supply chain need to be altered within a few months (see e.g. [15]). Automotive manufacturers and other players in the industry are called upon to change their product portfolios, due to technological change on the product side. This results from the increasing electrification of drive systems and the growing digitalization in the context of vehicle connectivity and automation. In addition, the entire industry has been confronted for some time now with the structural challenges that have developed globally over time. Complex supply chains, interdependencies within supply chains or susceptibility to failure due to lack of redundancy can be added as examples [16].

In the course of the corona pandemic, the susceptibility of value-added structures to disruption became particularly apparent [16]. The effects of limited supply and the resulting narrow value-added chains combined with the supply strategies commonly used in industry, such as just-in-time or just-in-sequence, led to major problems in companies' production networks. Furthermore, the pandemic also yielded to a crisis on the customer side. For example, vehicles could not be registered and brought to market as planned. Thus, existing but in the automotive industry no longer or only rudimentarily used approaches were moved into the focus of attention. For example, the regionalization of value creation, supplier management based on a wider number of suppliers, or the aspect of closed raw materials and material cycles were identified as possible ways out of the crisis [16–18]. In addition, there were calls for more proactive approaches. In the course of the Corona Pandemic it is not only about regaining the current state but much more important to develop approaches for the design of the "New Normal" [16]. In the face of future external uncertainties, the question now is how to design and build resilient supply and value systems with a stronger domestic focus.

3 Materials and Methods

There are many different approaches to the description and planning of value chains and networks. In the context of the distribution of value added to different partners or locations within a production network. Fleischer et al. [19] propose a planning approach. They introduce the term "value added module", which can be understood as a unit of several value-added activities with defined characteristics, such as freedom from overlap or a defined initial and final state. This concept can contribute both in the decomposition of value chains and in the associated analysis and definition of necessary resources. Furthermore, it can be used to determine the scope of value added in the context of new products [20]. In existing work, the modules were formed one-dimensionally according to defined criteria. The question now arises how modules should be built, considering permanent changes over time, like we experience them currently.

Manufacturing successful products is nowadays characterised by numerous interfaces and interactions with upstream and downstream processes and phases. Life cycle engineering can provide support by considering complex systems and their temporal dependencies from a system-oriented perspective. It thus forms a valuable extension of the previously described value-added module approach. Helpful descriptions of the different product phases for technical products and services can be found in the work of Spur and Krause [21], Meiren and Barth [22] or Block et al. [23] for example.

Considering value chains as a system under temporal change might contribute to finding long-lasting or permanent value-added solutions: Direct dependencies exist between different life cycle phases and between different elements of the value chain. For example, properties regarding producibility or functions for the utilisation phase following production must be considered and implemented already during the development of many new products (see Design for X [24]). This may also change the suppliers and their relationship within the production networks. While, for example, the development and integration of hardware takes place at a defined point in time within the development and production process, the corresponding software can be installed during the utilisation phase, if the appropriate conditions are met. As a result, supply chains may well change over the life cycle of a product. An anticipatory design of value creation therefore enables a flexible reaction to disruptions, like technological change or pandemic interruptions. Thus, strategic foresight approaches can make valuable contributions for a more proactive approach, as opposed to the reactive approach to unforeseen events which is still widespread in industry. By developing scenarios, external drivers of change in terms of technology, market or regulatory issues can be systematically analysed and extreme events and their effects simulated [25]. This, in turn, makes it possible to anticipate various alternative developments and react quickly to them.

4 Approach

Our methodology is based on linking the linear value-added module perspective with the product life cycle concept and strategic foresight approaches. The life cycle perspective considers temporal changes and dependencies between the individual supply chains of the life cycle phases. The value creation system is decomposed into manageable elements and the corresponding dependencies over selected phases of the life cycle are described. Herewith, two main questions are addressed. On the one hand, the question of how value chains can be subdivided or modularised in order to be as resilient as possible in the face of uncertainty and possible future changes (Step 1: Resilience-oriented modularisation of value-added). On the other hand, there is the question of how dependencies can be identified over different life cycle phases (Step 2: Resilience-oriented identification of dependencies). In order to answer these questions, a structural model (see Fig. 1) and a related procedure model (see Fig. 2) are described in the following. The necessary model elements are shown in Table 1.

The structural model describes the relationships between the individual model elements and their interactions. The model's core are alternative, future scenarios. These represent various possibilities for future value creation. On this basis, differences for value added can result in changes both within the product life cycle perspective and within the value chain perspective. Depending on the scenario, the life cycle phases of a product can differ for example in their characteristics. Furthermore, this can result in changes in the value-added chains within the individual phases

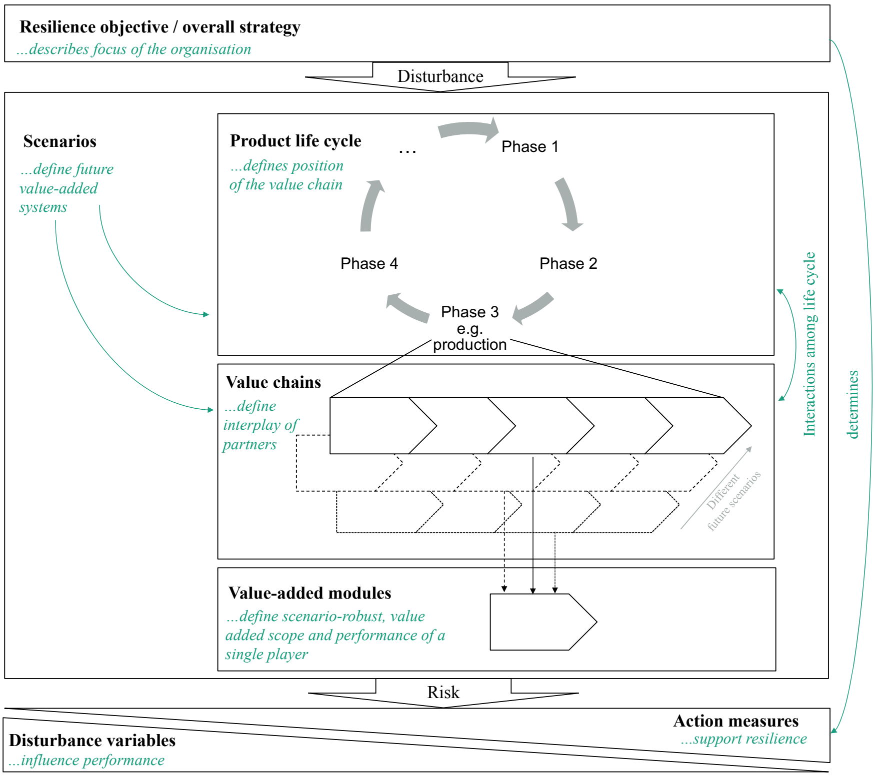


Fig. 1 Structure model for the description of the dependencies of the model elements

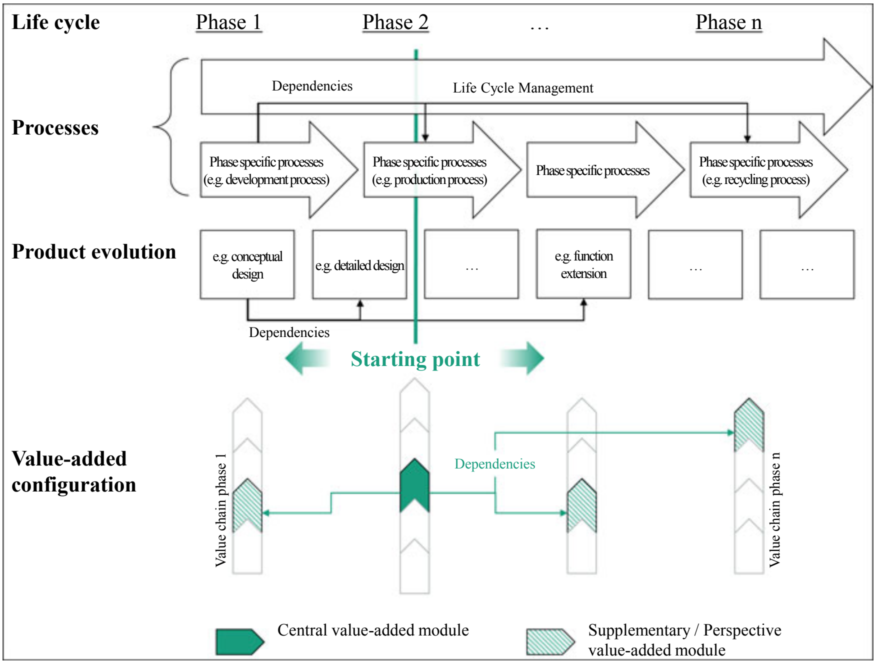


Fig. 2 Procedure for the identification of dependencies within specific processes and product evolution activities

of the lifecycle and guides the generation of scenario-robust value-added modules. Uncertainties regarding the achievement of objectives serve as input variables for generating the scenarios. The associated impacts are included in the analysis as risks and in turn determine the action measures to be taken. The disturbance variables and action measures are in turn dependent on the respective scenario.

The definition of scenario-robust value-added modules requires consideration of both perspectives, the value chain and the product life cycle which are directly linked to the process steps (1) and (2) described above. As a starting point for the analysis, a single life cycle phase with different value chain alternatives (one per scenario) is selected. The various scenarios and their effects on the value chains are examined. Interesting value-added activities are identified. Identification can be supported using different assessment approaches, for example, considering resources or competence criteria. In addition, depending on the degree of novelty of the product, a selection can also be made at the product level by means of differentiating features [20]. Subsequently, they are bundled into value-added modules, to form coherent and for the organisation attractive value-added activities that are robust across the various scenarios. Appropriate criteria can be used in the context of module formation. These can, for example, address aspects such as process logic and testability, technological or spatial coherence and temporal sequence [20]. Depending on the value chain's complexity of dimensions and the characteristics of the organisation, the frame of

Table 1 Overview of structural model elements

Element	Definition	Examples
Product life cycle	All phases a product goes through from completion to withdrawal from the market	E.g. product creation, product market, product disposal
Value chain	Process scope, which describes a defined section of the value creation performance of one or more value creation partners	E.g. multi-level supplier network for a drivetrain component within the automotive industry
Value-added module	Coherent value-added element within a value chain, defined by corresponding criteria	E.g. production of electrodes for battery cell manufacturing
Resilience objectives	Overarching objective to which the resilience efforts are geared	E.g. flexibility, robustness, responsiveness
Disturbance variable	Variable that influences the system in a limitedly predicable or unforeseen way	E.g. (natural) disasters, political unrest
Action measures	Action measures which are used to achieve the objective	E.g. site selection, sourcing and resource strategy, deployment of digital tools, digitalisation of process and product data
Scenario	Alternative vision of the future based on key factors and corresponding projections	E.g. scenarios for future mobility; dominance of battery electric drive systems versus hydrogen-based mobility

reference can be further limited to specific value-added activities within the value chain. It should be chosen in such a way that it is relevant for a player since it is either already part of his entrepreneurial activities today or will be in the future. Furthermore, it should also be reviewed whether upstream or downstream activities are integrated or, if necessary, outsourced to other partners.

Once one or more value-added modules have been identified for a life cycle phase, the next step is to consider the dependencies on other life cycle phases (see Fig. 2). It must be checked whether activities already carried out by the player within other life cycle phases need to be adapted due to the re-definition of one or more value-added modules in the initially considered life cycle phase. Such adaption dependencies are identified either through dependencies of processes or dependencies within product evolution among different life cycle phases. In addition to a possible adjustment, it may also be that new activities are required to generate future added value or that completely new value creation potential is created by the changed conditions. If significant potential for new value creation is identified in another life cycle phase, the process can be restarted in this life cycle phase. Potential value-added modules are derived which would lead to an expansion of the current scope of value added. In the following we present the concept using the example of automotive software and electronics.

5 Case Study: Automotive Software and Electronics

We will apply our framework to a use case within the automotive industry, to proof that it is suitable to identify interdependencies between value chains of multiple life cycle phases. A possible, future change in functionality of an electric vehicle serves as an example: The battery and charging system of an electric vehicle is planned to be extended after the vehicle is already on the market: It should support bidirectional instead of only unidirectional charging (see e.g., [26]). As such, the value chains in all life cycle phases need to be prepared and resilient to such changes. Our approach generates such supply chains over all life cycle phases.

The starting point of our framework (see Fig. 2) is the ‘usage’ phase in which the extension should take place. However, we look from it in an anticipating manner, as the vehicle is still in development. As such, the supply chain will also be prepared in an anticipating manner via our approach.

In a first step, potential future function attributes are estimated, based on scenario analysis and market research (see e.g., [27, 28]). Subsequently, technical implications for the charging function are derived. In the exemplary case, they point to increased communication and coordination needs between the battery management system and the vehicle charging control unit as well as between the vehicle charging control unit and the charging station. Decoupling of software and hardware allows to implement the necessary functional adaptions in a reactive manner in software, whenever necessary [29]. Possible alternations in hardware are done in an anticipating manner to save high adaption costs later [30]. However, these technical changes have effects on the supply chain and thus on the value-added modularization: Firstly, the initial design of the battery and charging system hardware have changed. Thus, the associated activities in the value-added module of the lifecycle phase ‘production’ have changed, due to an anticipating product design. The definition of the value-added module must be adapted (see Sect. 4). The supplier increases the resilience of its core activity in the ‘production’ phase, to support the creation of value in future life cycle phases for the manufacturer. From the supplier’s perspective, the extension of its activities means a technological leap on the one hand, while securing its further existence.

Secondly, the value-added module of integrated software and hardware development might be split up. Future adaptions in software and communication must be integrated into the same value-added module, to obey to the freedom from overlap principle (see Sect. 3). The vehicle charging control unit, the battery and charging hardware as well as the battery management system are currently defined as one value-added module. The supplier of this module must now provide an extension option for bidirectional charging to its value-added module for some time in the future. This directly affects current contract design: The supplier must provide the extension option and the manufacturer might pay a surcharge in exchange. However, the original assumptions under which the modularization happened, does not hold anymore; for example, if the supplier lacks the experience to implement bidirectional charging standards (e.g. [31, 32]). Externalizing only the software extension for bidirectional charging as a separate value-added module in the 'usage' phase violates freedom from overlap. Thus, the original value-added module is split in advance. As such, two different value-added modules exist: (1) Hardware development and production for the battery system and charging and (2) software development for uni- and bidirectional charge control. The supplier must now decide whether to subcontract the software development to still supply the whole value-added module or to lose activity parts in the value chain 'production'. Depending on the technological complexity, one or the other choice might be preferred. However, from the automotive manufacturer's perspective, the modules will now stay stable throughout possible changes in the lifecycle. The supply chain is robust against possible requirements for bidirectional charging in the future. Costs for resilience in this case include: Organizational costs for extended contract and development design as well as the option surcharge.

However, the adaption in the supply chain of the 'production' and 'usage' phases effects the shape of the supply chain in development. The development toolchain for software and hardware is a central integration point for different suppliers and thus defines the shape of the supply chain in development. If suppliers or modules change, the toolchain must be flexible enough to adapt with little effort. In our case this means, that software extensions, especially for bidirectional charging, must be supported in an anticipatory manner; for example, within the testing suits and hardware-in-the-loop testbeds. As such, the redefinition of the battery and charging value added modules has implications on the suppliers of these systems and the manufacturer's supply chain in all life cycle phase. Our framework was able to reveal that in a structured and guided manner.

6 Discussion and Conclusion

Within this paper, a new approach for the description of value added under temporal change is presented. We combine linear value-added processes with a life cycle-oriented perspective. This makes it possible to create scenario-robust value-added scopes on the one hand and, on the other, to address their interdependencies with further life cycle phases. Application of the framework was demonstrated using a case study of modern automotive software and electronics. The results indicate that the methodology is suitable to frame and identify scenario-robust value-added modules and point towards interdependencies between value chains of multiple life cycle phases. This is particularly important in the context of new products and services as well as their associated value-added processes and systems.

However, more knowledge needs to be built up about the inner steps for the creation of resilient value-added modules. Fleischer et al. [19] as well as Herrmann [20] provide steps to build the value-added modules. Yet, they need to be adapted to fit our approach. Additionally, further fields of application are to be addressed in future work. Promising examples are the production of future drive and energy storage systems.

To summarize, our methodological approach provides a framework to identify future challenges for and dependencies among different supply chains in a product's lifecycle. It serves as a first step towards an anticipating and proactive approach to generate resilient value-added modules. This is particularly relevant in the wake of disruptive environmental changes, like technological process or a global pandemic, to ensure the survival of local and global value creation. It is intended to transfer elements of the described framework into an executable simulation tool.

References

1. Tyler, S., Moench, M.: A framework for urban climate resilience. Clim. Dev. (2012). https://doi.org/10.1080/17565529.2012.745389
2. Holling, C.S.: Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. (1973). https://doi.org/10.1146/annurev.es.04.110173.000245
3. Pettit, T.J., Croxton, K.L., Fiksel, J.: The evolution of resilience in supply chain management: a retrospective on ensuring supply chain resilience. J. Bus. Logist. (2019). https://doi.org/10.1111/jbl.12202
4. Christopher, M., Peck, H.: Building the resilient supply chain. Int. J. Logist. Manag. (2004). https://doi.org/10.1108/09574090410700275
5. Sheffi, Y.: The Resilient Enterprise. Overcoming Vulnerability for Competitive Advantage, 1st edn. MIT Press, Cambridge (2007)
6. Wieland, A., Marcus Wallenburg, C.: The influence of relational competencies on supply chain resilience: a relational view. Int. J. Phys. Distrib. Logist. Manag. (2013). https://doi.org/10.1108/IJPDLM-08-2012-0243
7. Brandon-Jones, E., Squire, B., Autry, C., Petersen, K.J.: A contingent resource-based perspective of supply chain resilience and robustness. J. Supply Chain Manag. (2014). https://doi.org/10.1111/jscm.12050
8. Akgün, A.E., Keskin, H.: Organisational resilience capacity and firm product innovativeness and performance. Int. J. Prod. Res. (2014). https://doi.org/10.1080/00207543.2014.910624
9. Pizzol, M.: Life cycle assessment and the resilience of product systems. J. Ind. Ecol. (2015). https://doi.org/10.1111/jiec.12254
10. Scholten, K., Schilder, S.: The role of collaboration in supply chain resilience. Supply Chain Manag. (2015). https://doi.org/10.1108/SCM-11-2014-0386
11. Ponomarov, S.Y., Holcomb, M.C.: Understanding the concept of supply chain resilience. Int. J. Logist. Manag. (2009). https://doi.org/10.1108/09574090910954873
12. Fiksel, J.: Sustainability and resilience: toward a systems approach. Sustain. Sci. Pract. Policy (2006). https://doi.org/10.1080/15487733.2006.11907980
13. Kagermann, H., Süssenguth, F., Körner, J., Liepold, A., Behrens, J.H.: Resilienz der Fahrzeugindustrie: Zwischen globalen Strukturen und lokalen Herausforderungen. acatech IMPULS, München (2021)
14. Prawitz, S.: Studie: Versorgung mit Halbleitern schlechter als angenommen. Halbleitermangel. Vogel Automobil Industrie, 14 May 2021. https://www.automobil-industrie.vogel.de/studie-versorgung-mit-halbleitern-schlechter-als-angenommen-a-1024073/. Accessed 22 May 2021
15. teissOnline: Security flaw in ARMv7 allows hackers to gain control over smart cars (2020). https://www.teiss.co.uk/smart-cars-armv7-vulnerability/. Accessed 26 Nov 2020
16. Ishida, S.: Perspectives on supply chain management in a pandemic and the post-COVID-19 era. IEEE Eng. Manag. Rev. (2020). https://doi.org/10.1109/EMR.2020.3016350

17. Nationale Plattform Zukunft der Mobilität NPM: Qualitative Betrachtung des Wertschöpfungsnetzwerks Batterierecycling. Positionspapier (2020)
18. Nationale Plattform Zukunft der Mobilität NPM: Elektrische Maschine—Status Quo, Ausblick und Handlungsbedarf für die deutsche Wirtschaft. Positionspapier (2020)
19. Fleischer, J., Herm, M., Schell, M.-O.: Wertschöpfung in Netzwerken. ZWF (2004). https://doi.org/10.3139/104.100807
20. Herrmann, F.: Ein Verfahren zur Bestimmung des Wertschöpfungsumfangs bei neuen Produkten. Schriftenreihe zu Arbeitswissenschaft und Technologiemanagement, Band 27. Fraunhofer Verlag, Stuttgart (2015)
21. Spur, G., Krause, F.-L.: Das virtuelle Produkt. Management der CAD-Technik. C. Hanser, München, Wien (1997)
22. Meiren, T., Barth, T.: Service Engineering in Unternehmen umsetzen. Leitfaden für die Entwicklung von Dienstleistungen. Fraunhofer-IRB-Verlag, Stuttgart (2003)
23. Block, L., Riedel, O., Herrmann, F.: A lifecycle model to support continuous component evolution in embedded automotive systems. In: Bargende, M., Reuss, H.-C., Wagner, A., Wiedemann, J. (eds.) 19. Internationales Stuttgarter Symposium. Proceedings, pp. 1175–1189. Springer Fachmedien Wiesbaden, Wiesbaden (2019)
24. Luft, T., Wartzack, S.: Klassifikation und Handhabung von Unsicherheiten zur entwicklungsbegleitenden Erfassung des Produktreifegrades. In: Stelzer, R. (ed.) Entwerfen Entwickeln Erleben 2014. Beiträge zur virtuellen Produktentwicklung und Konstruktionstechnik, Dresden, 26–27 June 2014, pp. 535–549. TUDpress, Dresden (2014)
25. Gausemeier, J., Fink, A., Schlake, O.: Szenario-Management. Planen und Fuehren mit Szenarien, 2nd edn. Hanser, München (1996)
26. Menzel, S.: VW-News. Volkswagen will E-Autos als Energiespeicher nutzen. Handelsblatt, 5 Apr 2021. https://www.handelsblatt.com/mobilitaet/elektromobilitaet/elektromobilitaet-bidirektionales-laden-so-will-volkswagen-am-speichern-von-strom-verdienen/27052182.html. Accessed 19 May 2021
27. Bauer, W., Riedel, O., Herrmann, F., Borrmann, D., Sachs, C., Schmid, S., Klötzke, M.: ELAB 2.0. Wirkungen der Fahrzeugelektrifizierung auf die Beschäftigung am Standort Deutschland, Stuttgart (2019)
28. Stegmüller, S., Werner, M., Kern, M., Birzle-Harder, B., Götz, K., Stein, M.: Akzeptanzstudie Robocab. Autonome Mobilitätskonzepte aus Sicht der Nutzer, Stuttgart (2019)
29. Schäuffele, J., Zurawka, T.: Automotive Software Engineering. Grundlagen, Prozesse, Methoden und Werkzeuge effizient einsetzen, 6th edn. ATZ/MTZ-Fachbuch. Springer Vieweg, Wiesbaden (2016)
30. Maiyalagan, T., Elumalai, P.: Rechargeable Lithium-Ion Batteries. Trends and Progress in Electric Vehicles. CRC Press, Boca Raton (2020)
31. CHAdeMO Association: CHAdeMO Protocol. Revision 3.0(3.0) (2020)
32. International Organization for Standardization: Road Vehicles—Vehicle to Grid Communication Interface. Part 1: General Information and Use-Case Definition (15118-1) (2019)

Theory of Agency in Supply Chain Finance: Taking a Hermeneutic Approach

Luca Mattia Gelsomino, Rob Dekkers, Ronald de Boer, Christiaan de Goeij, and Qijun Zhou

Abstract This article investigates theory of agency as a theoretical underpinning in the field of supply chain finance. Specifically, through a hermeneutic approach, the authors examine the development of the theory, its postulations and assumptions, and its use in the field of supply chain finance. This leads to three conclusions: the theory is currently adopted rather superficially; there is potential for further developments by investigating non-standard configurations; and there is an increasingly relevant set of articles in supply chain finance that seems to be positioned within the boundaries of the theory of agency but makes no mention of it.

Keywords Agency theory · Supply chain finance · Hermeneutic approach

1 Introduction

There is little doubt that the topic of supply chain finance has gained considerable attention both among practitioners and researchers. Focusing specifically on the latter, there is ample evidence of the relevance that the topic has acquired in purchasing and supply management and supply chain management in general (evidenced, for example, by the recent special issues on the topic published in 2019 by the Journal of Purchasing and Supply Management and the forthcoming one by the International Journal of Operations and Production Management). Within academic literature focusing on supply chain finance, calls for strengthening the theoretical background of the topic have been multiple and repeated over time [1–3]. Despite these calls, it is manifest that developments in the theoretical foundation of supply chain finance,

L. M. Gelsomino (✉) · R. de Boer · C. de Goeij
Windesheim University of Applied Sciences, Zwolle, The Netherlands
e-mail: lm.gelsomino@windesheim.nl

R. Dekkers
University of Glasgow, Glasgow, Scotland

Q. Zhou
University of Greenwich, London, England

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_30

simply put, lurk. As a simple but powerful image, one could simply notice that several of the recent publications, despite clear relevance for supply chain finance, seem to lack any significant theoretical grounding [as explicitly noted by 3]. On the long run, such an approach will surely be detrimental to advancing scholarly insight into supply chain finance, limiting and inhibiting proper scientific foundations, which restrict knowledge transfer, development of practice and innovation through incoherent understanding of supply chain finance and its instruments. Taking these considerations, the discourse described in this paper focuses specifically on one theoretical underpinning, namely the theory of agency, which seems to be among the ones that provide the best explanatory power for the field of supply chain finance [4].

Three literature reviews focusing on supply chain finance [1–3], plus one focusing specifically on sustainable supply chain finance practices [5], have already been published, while several other reviews on topics partially overlapping with supply chain finance can be found in literature [e.g. 6]. Thus, this contribution does not aim at performing a further literature review but, against the backdrop of these and other existing contributions, poses the following questions: (i) how the theory of agency is explicitly or implicitly used in the supply chain finance discourse; (ii) what are any potential confusions or calls for additional scientific accuracy, and finally (iii) how future contributions in this area can advance by using the theory of agency.

The authors of this paper, while being committed and engaged in the supply chain finance discourse, share the concern that confusion around the use of the theory of agency will erode scientific trustworthiness and rigorous in future publications, with missed opportunities for understanding of phenomena and applications for supply chain finance, and consequently lack of influence on innovative practices. With such concerns in mind, this paper describes how we used philosophical hermeneutics to interpret relevant literature and come to a common understanding and conclusions for improving the trustworthiness of studies that wish to theoretically ground their contribution in the theory of agency. The writing starts by illustrating our method (philosophical hermeneutics). It then proceeds to briefly summarising the theory of agency, focusing on its position within the field of supply chain finance, how we developed our research question, the processes that we undertook to address this question and the answers that we discovered. We discuss our interpretations with respect to the use of the theory of agency within the field of supply chain finance and end with recommendations for those who engage in research within this field of enquiry.

2 Method

As mentioned above, this article employs philosophical hermeneutics as a research method, as appropriate for interpretation of complex or conflicting issues in interpreting existing texts. Specifically, we rely on the work of Gadamer [7] and Shephard et al. [8]. While aiming at reaching a common understanding among the author team, there are two fundamental aspects of philosophical hermeneutic enquiry, as illustrated by Shephard et al. [8] and Gadamer [7, pp. 295, 403]. The first is an interpretation of text analysis that relates reading a text to having a conversation with it, so as to reach an understanding. The second describes the inevitability of interpretation being the product of productive or unproductive prejudices.

The early discussion leading to this article was initiated by a subset of the author team, as part of a generalised conversation on the topic of supply chain finance as a scientific discipline, as a consequence of an earlier study [9] accepted for publication at the time. This led the authors to investigate the theoretical foundations of the topic, adopting a comprehensive approach aimed at broadly evaluate all possible theoretical underpinning within the topic. After an initial phase focused on a general examination of theoretical conceptualisations around supply chain finance, we decided to focus specifically on the theory of agency. This was accelerated by a presentation for a special interest group [10] in which a particular finding with regard to the theory of agency was further elaborated based on the earlier study. This led to a year of work in which the author team regularly met, exchanging insights and notes on this specific topic through videoconferencing as well as emailing. In discussion ideas and notes on the topic were shared and new and additional papers on supply chain finance, theory of agency or any other relevant topics were shared for further reading. Once questions arose in previous meetings were resolved, the new understanding of the topic was shared in notes and write-ups, and further readings were shared and discussed in subsequent meetings, re-starting the cycle. Eventually, after several rounds of discussion and analysis, we formulated the following research question: *how does theory of agency implicitly or explicitly affect the conceptualisation and development of supply chain finance?*

3 Theory of Agency and Its Use in Supply Chain Finance

Once chosen to focus on theory of agency in supply chain finance, we first decided to analyse the three publications that are commonly referred as the basis for theory of agency, namely: Mitnick [11], Ross [12], and Jensen and Meckling [13]. It appeared that the theory of agency describes risk sharing by dependent actors in a social-economic relationship based on three principles, which can be derived from the three original sources. The first is that a principal delegates activities to an agent with the purpose to act on its behalf [11, pp. 1, 6; 12, p. 134; 13, p. 308]. The second principle is that there is an optimum for the allocation of resources, both in terms of performing the activity and its governance [12, p. 134; 13, p. 309]. The third principle is that access to information for finding this optimum is perfect for both the principal and agent, while sharing this information between themselves [11, p. 6; 12, p. 135]. All three sources [11, p. 13; 12, p. 138; 13, p. 308] indicate that the third principle may unlikely hold true, giving rise to the '*principal-agent problem*' [11, p. 14].

Despite the wide-ranging body of literature developed with direct reference to the theory of agency, it was surprising for us to realise that only two significant extensions from the original concept could be identified. The first, effectively summarised by Fayezi et al. [14], points to the extensive use of the so-called positive theory of agency within the domain of supply chain management. Explicitly coined by Jensen [15], the objective function can be considered normative, whereas the valuation function based on the availability of alternatives is positive. The second development emerged when reading Panda and Leepsa [16], who also refer to positive agency theory, and delineate it as explaining causes for the agency problem and costs associated with it. Furthermore, they [16, p. 79] point out that criticism of the positive agency theory has led to the emergence of the behavioural agency theory. This development for theory of agency focuses more on the agent's behaviour than solely on the objective of the principal. These extensions mean that three variations for the theory of agency exist: theory of agency as the principal-agent problem, positive theory of agency and behavioural agency theory.

After having eviscerated the history of the theory of agency and its development, we then felt comfortable returning to the analysis of supply chain finance. We decided to use the classical framework of postulations and underlying assumptions, presented by Gouldner [17]. Two of us did already have experience with this for another, unrelated study. We mapped out the underlying assumptions for postulations, divided into domain and background assumptions. These 'implicit' assumptions together with the 'explicit' ones allowed us to further interpret and highlight potential shortcomings in the use of the theory in recent literature. Due to space restrictions, we cannot present more details on those specific assumptions, or the process that led to their discovery.

In analysing various articles on supply chain finance that mentioned theory of agency, it became rapidly clear that a large group mentioned it *en passant*, without any apparent implication for their contribution. Examples include Liebl et al. [18] or Wetzel and Hofmann [19]. Despite claims that the theory of agency is one of the most widely used theories in supply chain finance, only a relatively minor and limited set of papers is actually using it as the theoretical underpinning of the contribution. More importantly, out of those papers we could only find three that actually go beyond applying constructs from the theory '*as is*' and investigate in-depth the implications emanating from analysing supply chain finance through the lens of the theory of agency. Specifically, Martin and Hofmann [20] highlight two instances of the principal-agent problem in the analysed cases for supply chain finance: between buyer and supplier and within the same company, between the finance (principal) and procurement and operations (agent), casting a light on the relevance that the theory might have for different 'principal-agent constellations.' Wandfluh et al. [21] identify a triad composed of the financial and purchasing departments in a buyer organisation (both principals) and the supplier (agent), expanding on literature focusing on the opportunistic opportunities for agents when multiple principals lack alignment. More marginally, Ali et al. [22, p. 3] sees the financial institution as the principal and the buyer as the agent, opposed to the mainstream view that investigates the buyer as principle and the supplier as agent. This first iteration led us to our first conclusion:

> The theory of agency tends to be applied sparingly and superficially in supply chain finance, causing risk of flawed research design and missed opportunities in research in this field.

This has clear implications. For example, scrutinising the aforementioned studies highlighted the relevance of applying the theory not only at the level of the organisation, but also to relationships between specific departments within firms involved. This resonated with us, because it aligned with findings from the Scottish focus group. Not doing so might cause problems, for instance when administrating a questionnaire. The identifications of (multiple) triads as relevant constructs for supply chain finance is equally relevant for future developments. However, a superficial understanding of the theory, which seems apparent in several contributions, is likely to taint the study design, limiting the possibility of adopting a triad as a unit of analysis from the very beginning.

Starting Martin and Hofmann's [23] remark regarding principal-agent configurations, we then investigated potential non-standard configurations for principal-agent roles, to understand if other authors had picked this up or investigated it in parallel. We first looked into the entire literature, without focusing on supply chain finance, and we found papers that were mentioning the 'reverse' of roles between principal and agent [e.g. 24, 25]. Focusing on supply chain finance, only Dekkers et al. [4, p. 10] mentioned the notion of the reversal of roles when suppliers act as principal rather than agent. However, the inversion of roles in the principal-agent relationship, and its implications, are especially evident when analysing supply chain finance practices. As example serves the practice of reverse factoring. In it, a buyer authorises a financial institution to pay to its suppliers specific ('approved') invoices ahead of their contractual tenure. The financial institution will then charge the supplier a fee for the service and the buyer will repay the financial institution when the original contractual payment date is due [e.g. 26]. It is straightforward to examine this arrangement through the lens of theory of agency, with the supplier as the principal, delegating the action of 'approving' an invoice and authorising the financing act to the buyer (the agent), asking the latter to act on his behalf with the financial institution. However, streaming directly from the first conclusion, this does not seem to happen, as the theory is adopted often superficially. This led to our second conclusion:

> Adopting the theory of agency as theoretical lens with an inversion of roles between buyer and supplier might provide new opportunities for advancing the understanding of the principal-agent problem in supply chain finance and supply chain management in general.

Up to this point we limited our discussion to articles that were explicitly mentioning theory of agency. However, the search for papers that inversed the role of buyer and suppliers led us to realise that several contributions within the field of supply chain finance, especially from an analytical perspective, adopted some or all of the postulations and assumptions of the theory but did not mention it explicitly. Examples include Gelsomino et al. [27], Huang et al. [28] and Cai et al. [29]. All three papers do not explicitly discuss any theoretical background for their contribution, other than analysing existing literature and employing common modelling techniques. However, all of them implicitly present postulations that are typical of the theory of agency (e.g. the existence of an optimal fee structure to find an equilibrium between a buyer and a supplier, or the fact that the principal has all information required for a rational decision). The lack of theoretical grounding is detrimental as it limits structured developments in the supply chain finance field. In this case, it limits the ability of modelling contributions to effectively challenge existing explicit or implicit assumptions in both the theory (e.g., fixed roles) and in the field under examination (e.g., are buyers always acting rational when offering supply chain finance to their suppliers? What are the consequences if they are not?). This leads to our third and final conclusion:

> Modelling contributions in supply chain finance should be more explicit in their theoretical underpinning, creating opportunities to challenge implicit or explicit assumptions in both the theory of agency and supply chain finance as a field.

4 Conclusions

This article has investigated the use of the theory of agency in the field of supply chain finance, through a hermeneutic approach. This process led us to three conclusions, highlighting a certain widespread superficial use of the theory, the potential to investigate non-standard configurations and a subset of modelling papers that seem to sit within the boundaries of the theory but make no mention of it. Whereas we more or less expect the first conclusion, the scrutinising of a wide range of papers, not limited to the ones mentioned in this paper, we did not expect our journey to find the latter two in later stages of the development of our thinking.

These three conclusions have implications for researchers, which are explained in the section above. However, they also bear clear, albeit indirect, implications for practice. As we illustrated in the introduction, sub-optimal use of the theory of agency is likely to cause an oversight of opportunities for development of supply chain finance, which will have an impact on the potential for academia to influence and foster innovation practices in this field. We see this conversational contribution as a call for a more critical use of the theory and, more importantly, a critical look at its explicit and implicit assumptions in supply chain finance and supply chain management.

References

1. Gelsomino, L.M., Mangiaracina, R., Perego, A., Tumino, A.: Supply chain finance: a literature review. Int. J. Phys. Distrib. Logist. Manag. **46**, 348–366 (2016). https://doi.org/10.1108/IJPDLM-08-2014-0173
2. Xu, X., Chen, X., Jia, F., Brown, S., Gong, Y., Xu, Y.: Supply chain finance: a systematic literature review and bibliometric analysis. Int. J. Prod. Econ. **204**, 160–173 (2018). https://doi.org/10.1016/j.ijpe.2018.08.003
3. Jia, F., Blome, C., Sun, H., Yang, Y., Zhi, B.: Towards an integrated conceptual framework of supply chain finance: an information processing perspective. Int. J. Prod. Econ. **219**, 18–30 (2020). https://doi.org/10.1016/j.ijpe.2019.05.013
4. Dekkers, R., de Boer, R., Gelsomino, L.M., de Goeij, C., Steeman, M., Zhou, Q., Sinclair, S., Souter, V.: Evaluating theoretical conceptualisations for supply chain and finance integration:

a Scottish focus group. Int. J. Prod. Econ. **220**, 107451 (2020). https://doi.org/10.1016/j.ijpe.2019.07.024
5. Jia, F., Zhang, T., Chen, L.: Sustainable supply chain finance: towards a research agenda (2020)
6. Seifert, D., Seifert, R.W., Protopappa-Sieke, M.: A review of trade credit literature: opportunities for research in operations. Eur. J. Oper. Res. **231**, 245–256 (2013). https://doi.org/10.1016/j.ejor.2013.03.016
7. Gadamer, H.-G.: Truth and Method. Continuum, London (2004)
8. Shephard, K., Rieckmann, M., Barth, M.: Seeking sustainability competence and capability in the ESD and HESD literature: an international philosophical hermeneutic analysis. Environ. Educ. Res. **25**, 532–547 (2019). https://doi.org/10.1080/13504622.2018.1490947
9. Dekkers, R., Kühnle, H.: Appraising interdisciplinary contributions to theory for collaborative (manufacturing) networks: still a long way to go? J. Manuf. Technol. Manag. **23**, 1090–1128 (2012). https://doi.org/10.1108/17410381211276899
10. Dekkers, R., de Boer, R., Gelsomino, L.M., de Goeij, C., Steeman, M., Zhou, Q.: Introducing the Inverse Principal-Agent Problem (2019)
11. Mitnick, B.M.: Fiduciary rationality and public policy: the theory of agency and some consequences. In: Proceedings of the American Political Science Association, 1973, pp. 1–40, New Orleans (1973)
12. Ross, S.A.: The economic theory of agency: the principal's problem. Am. Econ. Rev. **63**, 134–139 (1973). https://doi.org/10.2307/1817064
13. Jensen, M.C., Meckling, W.H.: Theory of the firm: managerial behavior, agency costs and ownership structure. J. Financ. Econ. **3**, 305–360 (1976). https://doi.org/10.1037/1082-989X.7.3.301
14. Fayezi, S., O'Loughlin, A., Zutshi, A.: Agency theory and supply chain management: a structured literature review. Supply Chain Manag. **17**, 556–570 (2012). https://doi.org/10.1108/13598541211258618
15. Jensen, M.C.: Organization theory and methodology. Account. Rev. **58**, 319–339 (1983). https://doi.org/10.2139/ssrn.94036
16. Panda, B., Leepsa, N.M.: Agency theory: review of theory and evidence on problems and perspectives. Indian J. Corp. Gov. **10**, 74–95 (2017). https://doi.org/10.1177/0974686217701467
17. Gouldner, A.W.: The Coming Crisis of Western Sociology. Heinemann, London (1971)
18. Liebl, J., Hartmann, E., Feisel, E.: Reverse factoring in the supply chain: objectives, antecedents and implementation barriers. Int. J. Phys. Distrib. Logist. Manag. **46**, 393–413 (2016). https://doi.org/10.1108/IJPDLM-08-2014-0171
19. Wetzel, P., Hofmann, E.: Supply chain finance, financial constraints and corporate performance: an explorative network analysis and future research agenda. Int. J. Prod. Econ. **216**, 364–383 (2019). https://doi.org/10.1016/j.ijpe.2019.07.001
20. Martin, J., Hofmann, E.: Towards a framework for supply chain finance for the supply side. J. Purch. Supply Manag. **25**, 157–171 (2019). https://doi.org/10.1016/j.pursup.2018.08.004
21. Wandfluh, M., Hofmann, E., Schoensleben, P.: Financing buyer–supplier dyads: an empirical analysis on financial collaboration in the supply chain. Int. J. Logist. Res. Appl. **19**, 200–217 (2016). https://doi.org/10.1080/13675567.2015.1065803
22. Ali, Z., Gongbing, B., Mehreen, A.: Supply chain network and information sharing effects of SMEs' credit quality on firm performance. J. Enterp. Inf. Manag. **32**, 714–734 (2019). https://doi.org/10.1108/JEIM-07-2018-0169
23. Martin, J., Hofmann, E.: Selecting financial service providers for supply chains: how cross-functional collaboration can improve effectiveness and efficiency. In: EUROMA—22nd Conference of the European Operations Management Association, pp. 1–10, Neuchâtel (2015)
24. Matthew, D.B.: Controlling the reverse agency costs of employment-based health insurance: of markets, courts, and a regulatory quagmire. Wake Forest Law Rev. **31**, 1037–1074 (1996)
25. Quick, S.W.: Staying afloat in the stream of commerce: Goodyear, McIntyre, and the ship of personal jurisdiction. North Carolina J. Int. Law Commer. Regul. **37**, 547–608 (2011)

26. van der Vliet, K., Reindorp, M.J., Fransoo, J.C.: The price of reverse factoring: financing rates vs. payment delays. Eur. J. Oper. Res. **242**, 842–853 (2015). https://doi.org/10.1016/j.ejor.2014.10.052
27. Gelsomino, L.M., de Boer, R., Steeman, M., Perego, A.: An optimisation strategy for concurrent supply chain finance schemes. J. Purch. Supply Manag. **25**, 185–196 (2019). https://doi.org/10.1016/j.pursup.2018.07.004
28. Huang, J., Yang, W., Tu, Y.: Financing mode decision in a supply chain with financial constraint. Int. J. Prod. Econ. **220** (2020). https://doi.org/10.1016/j.ijpe.2019.07.014
29. Cai, Y.-J., Chen, Y., Siqin, T., Choi, T.-M., Chung, S.-H.: Pay upfront or pay later? Fixed royal payment in sustainable fashion brand franchising. Int. J. Prod. Econ. **214**, 95–105 (2019). https://doi.org/10.1016/j.ijpe.2019.03.025

The Role of Machine Learning in Supply Chain Management

Thais Carreira Pfutzenreuter, Edson Pinheiro de Lima, Sergio Eduardo Gouvêa da Costa, and Fernando Deschamps

Abstract The world experienced historic challenges in global demand management as a result of the covid-19 pandemic. Supply chains were abruptly interrupted with new traffic rules among countries and organizations confronted hard moments of absolute uncertainty with an extreme complex planning scenario while customers demanded resources on time in order to stay safe at home during confinement. For this reason, supply chain risk management and demand forecasting with artificial intelligence has become even more explored by the scientific community. In this context, this paper proposes an investigation of machine-learning projects' contribution for supply chain management in organizations, not only during the pandemic crisis, but during the last recent years. PRISMA approach was applied for a systematic literature review, limiting English written articles indexed at Scopus and complementary sources, such as Science Direct and IEEE. Results have defended the increasingly important role of machine-learning projects in supporting organizations to plan their operational demands and activities, improving operational efficiency and strengthening strategic supplier selection even in challenging pandemic times. The main contribution is focused on examining theoretical relationships among recent approaches and address mutual strategic achievements through a diagram. Results presented by a summarized diagram exposed machine learning strategic value for demand forecasting, supply risk mitigation, lead-time reduction, greener operations and strategic supplier selection. Some common best practices observed revealed training and test segmentation, feature importance analysis and dimensionality reduction. Limitations are linked to further research suggestions, increasing the present bibliographic portfolio selection along with case-study implementations in order to extend connections between theory and practice.

T. C. Pfutzenreuter (✉) · E. P. de Lima · S. E. G. da Costa · F. Deschamps
Industrial and Systems Engineering, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil
e-mail: thais.pfut@gmail.com

E. P. de Lima · S. E. G. da Costa
Industrial and Systems Engineering, Universidade Tecnológica Federal do Paraná, Curitiba, Brazil

F. Deschamps
Mechanical Engineering, Universidade Federal do Paraná, Curitiba, Brazil

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_31

Keywords Machine learning · Artificial intelligence · Operations strategy · Supply chain management

1 Introduction

Managing risks in supply chains at a local or global scale has increasingly attracted attention from researchers and organizations in recent years. The field of supply chain risk management, which emerged in the early 2000s has now become more than the overlap of directly related areas such as enterprise risk management and supply chain management. With the increasement of big data analytics and artificial intelligence, machine learning projects have been implemented for supply chain risk management, operational performance assessments and improvements, greener operational workflows and systems and strategic supplier policies. The purpose of the present paper is to investigate the role of machine learning projects in supply chain management, sustained by a research design focused on applications from recent studies.

2 Theoretical Background

Machine learning is an interdisciplinary research area in the wider field of artificial intelligence, which reunites ideas from several branches of science, such as big data analytics, statistics and computational science. Machine learning is focused on developing fast and efficient learning algorithms that can make predictions based on historical data [3]. Therefore, machine learning is also known as predictive analytics, which aims to determine what is more likely to happen.

Machine learning projects have been increasingly applied by healthcare environment, providing accurate diagnosis, helping to predict mental health disease [30] or obstructive sleep [28]. Its multidisciplinary has also reached the educational setting [13], sport results predictions [24] and even law field of practice [7], with predictive judicial analytics identifying behavioral anomalies and feature relevance in judicial decision-making.

Some companies have applied artificial intelligence for customer retention [27], identifying potential customers based on purchase behavior [8] and even employee performance prediction through historical datasets [15]. In supply chain management environment, some works propose machine learning-based frameworks for managing the inventory at all nodes of the supply chain [21], implementing different algorithms for reinforcement learning to find near-optimal policies.

Prediction of lead time and cycle time key performance indicators are critical for good production planning programs, since they are used to be built based on the assessment of these metrics. Gyulai et al. [12] applied a lead time prediction based on regression algorithms for a real flow-shop environment exposed to frequent changes and uncertainties resulting from the changing customer order stream. Wang et al. [32] have also used a recurrent neural network for short-term cycle-time forecasting of wafer lots. Moreover, machine learning has been used to improve manufacturing processes. In this context, Diaz-Rozo et al. [10] proposed a cyber-physical system for machine component knowledge discovery based on clustering algorithms using real data from a machining process.

Demand volatility is also a challenging risk for supply chains. Hence, researchers and practitioners have focused on approaches for reducing unexpected demand volatility threats [9]. Therefore, machine learning has been a strategic resource to capture the behavior of volatile demand and diminish the uncertainty of policies replenishment [26].

3 Research Design

The research design followed the method outlined in the Preferred Reporting Items for Systematic review and Meta-Analysis, PRISMA Statement [20]. The PRISMA flow chart performed is illustrated by Fig. 1.

Machine learning and supply chain management themes were search with the reunion of 1st and 2nd axes linked by Boolean Operator "AND". The initial conjunction of Scopus and other sources, such as Science Direct, resulted in 234 papers after duplicates removal and 63 English written and open-accesses papers were restricted. The final filter excluded articles that did not address the intended conjunction target of both themes and bibliographic portfolio composition resulted in 20 papers after the selection of the studies that best addressed the research purpose.

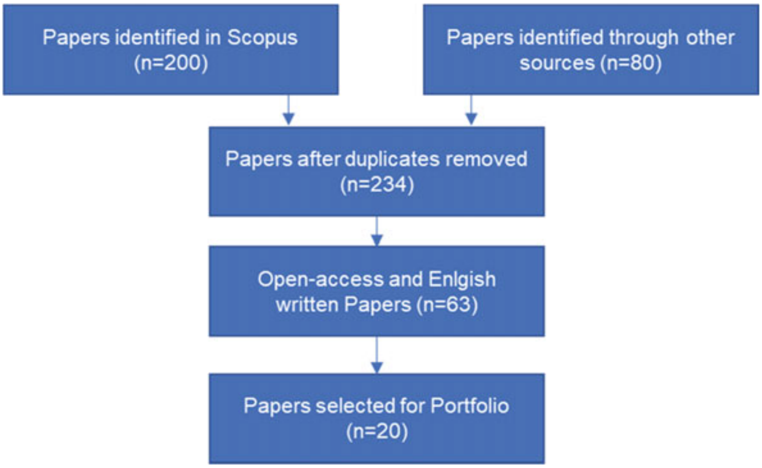


Fig. 1 Research design

4 Results

4.1 Strategic Supplier Selection

Efficient inventory management is extremely important for competitive financial performance and Priore et al. [26] argue that mapping unnecessary costs, assessing possible contract breaks, maintaining stable production schedules and smart inventory policies are some essential actions for this accomplishment. However, best practices are challenging in the routine of a productive system, since facing a dynamic and unstable scenario requires rapid changes in planning and many management techniques may not be sufficient for a good adequacy of stock policies in the long term. Machine-learning has contributed to improvements in supply chain management, based on demand forecasts and historical inventory flows. In this context, Priore et al. [26] went beyond the material flow data and proposed the construction of a dynamic structure with inductive learning, finding the most advantageous replacement policy over time, reacting to changes. Variables such as cost structure, demand variability, delivery terms and stock policies of different partners were used to build the forecasting model and, among four alternatives, the selection of the best stock policy presented 88% of accuracy, enabling considerable cost savings compared to the static supplier policy management previously used by the organization. The researchers observed that the gains were even greater when the external environment was more volatile than expected.

A smart supplier selection model was also developed by Cavalcante et al. [6] using supervised learning. Supply date and quantity of delivery data were strategically used as predictor variables to perform the supplier's reliability qualification, along with a simulation model that covered external, internal material flow and delivery to customers. Hybrid algorithms were performed with Random Forest, Linear Regression and K-Nearest Neighbors. The application of the combined approaches empowered risk mitigation strategies in supply chain management. Moreover, risk supplier mapping and the simulation of internal operations enabled cost reduction and operational speed with digital supplier management.

4.2 Demand Forecasting and Production Planning

Forecasts have traditionally served as the basis for planning and executing supply chain activities, driving smart decisions, shortening lead times and an improving resource management. With the big data era, recent research [14] reviews the impact brought by this explosion of data on product forecasting and how such data can provide insights for consumer behavior [4] and price optimization [11]. Xia et al. [34] proposed a hybrid method based on machine-learning with adaptive input metrics to improve the accuracy of sales forecasts and used real data from three sophisticated, medium and basic fashion retailers in Hong Kong to validate the suggested approach.

Bohanec et al. [3] complementarily explored the explanation of black box machine learning forecasting and propose an intelligent model in a real case of B2B sales forecasting.

The promotional impact on sales was explored by Abolghasemi et al. [1]. The researchers investigated 843 real demand time series to simulate promotion scenarios bringing volatility to demand scheduling programs and implemented some machine learning approaches for demand forecasts. ARIMA with covariate, support vector regression and dynamic linear regression presented good results for volatile demand forecasting in promotion scenarios. Pereira and Frazzon [25] had a similar research ambition, introducing a data-driven approach that combines machine-learning demand forecasting and operational planning simulation-based optimization to adaptively synchronize demand and supply in omni-channel retail supply chains. The authors applied a combination of clustering and neural networks improved demand forecast, supporting an assertive identification of demand volume and location, which reduced lead-time and operational costs. Moreover, the study succeeded on building a connection between digital and physical environments in retail supply chain.

Sales forecasting is also directly connected to more sophisticated production planning programs. Waschneck et al. [33] developed a real-time planning proposal based on reinforcement learning using multiple dispatch strategies and offline learning techniques. The approach applied Google DeepMind's Deep Q Network agent algorithm for Reinforcement Learning, achieving smart production control with deep neural networks and user-defined objectives to optimize scheduling. The proposed system was successfully validated in a small factory simulation.

In a more recent global scenario, a predictive analytics project has been recently developed by Nikolopoulos et al. [22] during the COVID-19 pandemic with the purpose to predict epidemiological growth rates with machine- and deep-learning models, together with a hybrid approach based on nearest neighbors and clustering. The authors used five distinct countries databases for the study, including United States, United Kingdom, Germany, Singapore and India and succeeded on helping planners for assertive decisions during the uncertain pandemic period implementing excess demand for products predictions using google trends data and simulations for lockdown and restriction policies.

4.3 Supply Risk Mitigation

Supply chain risk mitigation has been given attention for machine learning applications in industry, since disruptive moments can cause meaningful high costs. In a risk mitigation case study, Brintrup et al. [5] applied interruptions forecast in the first layer of supply chain in an equipment manufacturer's database. The training model required some challenges, once interruptions in industry represent a class imbalance where the sample of successful orders is much larger than interrupted orders. Hence, authors had to collect five samples per attribute for pre-processing, because with the excess of variables each attribute would become sparse, making statistical significance more difficult. Delivery and supply time were some of the predictor variables and Random Forest presented the best result. However, the authors [5] discussed that the interruption forecast reached a better remarkable performance when agility variable was inserted in the pre-processing models, introducing knowledge from the historical capacity of suppliers to handle greater monthly variations.

Ojha et al. [23] also defends supply chain risk propagation is a cascading effect of risks on global supply chains and proposes a holistic measurement approach based on Bayesian network for predicting the complex behavior of risk propagation and consequently improve supply risk management through the assessment of behavior risks.

Transportation disruption represents likewise a critical risk for supply chains, which affects many involved stakeholders. After transportation disruption, market demand becomes hard to predict, forcing enterprises to improve market demand and inventory predictions. In this context, Liu et al. [17] successfully implemented an improved model of grey neural networks algorithm to help enterprises with demand predictions after the occurrence of transport disruptions. Another similar transport risk approach was managed by a cargo logistics, where no deviation is expected between actual arrival time from planned arrival time. Neither earliness nor tardiness is desirable for customer and freight forwarders. Shang et al. [29] investigated ways to forecast transport risks using a half-year of air cargo data, provided by a leading forwarder on 1336 routes served by 20 airlines. Using a Bayesian nonparametric model, the authors provide a tool for the forwarder to offer customized price and service quotes, which additionally enables a fair supplier evaluation and tell apart recurrent risks from disruption risks.

4.4 Process and Operations

Strategic management of operations and processes is essential for a competitive supply chain management. Lot cycle time reduction in semiconductor manufacturing industry was explored by Meidan et al. [19], that investigated the identification of the most influential predictor variables using conditional mutual information maximization and sequentially applied the selective Naive Bayesian Classifier for selecting the minimal, most discriminative key-factor set for an accurate lot cycle time prediction. The achievement enabled better operational performance in industrial activities.

Efficient inbound logistics plays also an important role in operational performance, where processes depend on frequently changing information of products, assembly line planning and a high number of parts are dependent on different suppliers. Knoll et al. [16] contributed with an inbound logistics planning project supported by machine learning for internal process improvement. Even though inbound processes represent a small part in supply chain activities, its improvement brings cost reduction to the whole supply system and facilitates the work of planners and specialists.

Machine learning approaches have also been used as a tool for improving the operations of a third-part logistics. Tufano et al. [31] applied clustering algorithms with this purpose to define product families for the assignment of the workload processing recourse considering different product family scenarios and distinct tuning parameters. The impact of each clustering on operations is measured and assessed via logistic KPI dashboards.

4.5 Sustainability and Green Supply Chain

Big data analytics has contributed for mitigating social issues in the supply chain management. Mani et al. [18] concluded machine learning and big data have supported companies to predict various social problems including workforce safety, fuel consumptions monitoring, workforce health and security among many other issues, highlighting the importance of information management for social risk mitigation. Benzidia et al. [2] complementarily investigated the impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance. A conceptual model was developed to test a sample of data from 168 French hospitals using partial least squares regression-based structural equation. Results have shown the use of big data and artificial intelligence technologies has a significant effect on environmental process integration and green supply chain collaboration.

4.6 The Role of Machine Learning in SCM

Machine learning role in supply chain management is summarized by a diagram presented by Fig. 2. Performance measures including supply date, delivery volume, variability and supplier lead times are some of the commonly performance indicators collected by researchers and project managers to build machine learning predictions in supply chains.

The diagram summarizes mutual strategic outputs of recent projects, such as supply risk mitigation, lead-time reduction, greener systems and more reliable products and services, increasing customer satisfaction and enhancing strategic supplier selection along with green supply chain management. Common best practices highlighted by bibliographic portfolio composition revealed training and test segmentation, feature importance analysis and dimensionality reduction.

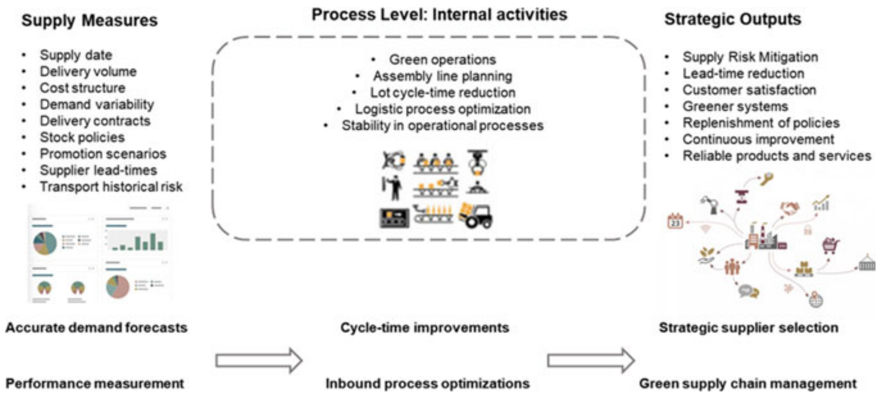


Fig. 2 Diagram of machine learning role in SCM

5 Conclusion

The main contribution of this paper consisted of strengthening machine learning contribution to strategic supply chain management, identifying important achievements by recent projects through different perspectives of demand forecasting, supply risk mitigation, operational processes and sustainability in supply chains. Results presented by a summarized diagram exposed machine learning strategic value for demand forecasting, lead-time reduction, greener systems and strategic supplier selection. Beyond these common achievements, training and test segmentation, feature importance analysis and dimensionality reduction were mutual approaches for high performing predictive analytics. We expect to cover some topics and limitations left unexplored by this initial study, extending bibliographic portfolio composition and applying case-study analysis to complement the present investigation, connecting theory to practice.

References

1. Abolghasemi, M., Beh, E., Tarr, G., Gerlach, R.: Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion. Comput. Ind. Eng. **142**, 106380 (2020). https://doi.org/10.1016/j.cie.2020.106380
2. Benzidia, S., Makaoui, N., Bentahar, O.: The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance. Technol. Forecast. Soc. Change **165**, 120557 (2021). https://doi.org/10.1016/j.techfore.2020.120557
3. Bohanec, M., Kljajić Borštnar, M., Robnik-Šikonja, M.: Explaining machine learning models in sales predictions. Expert Syst. Appl. **71**, 416–428 (2017). https://doi.org/10.1016/j.eswa.2016.11.010

4. Boone, T., Ganeshan, R., Jain, A., Sanders, N.R.: Forecasting sales in the supply chain: consumer analytics in the big data era. Int. J. Forecast. **35**(1), 170–180 (2019). https://doi.org/10.1016/j.ijforecast.2018.09.003
5. Brintrup, A., Pak, J., Ratiney, D., Pearce, T., Wichmann, P., Woodall, P., McFarlane, D.: Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. Int. J. Prod. Res. **58**(11), 3330–3341 (2020). https://doi.org/10.1080/00207543.2019.1685705
6. Cavalcante, I.M., Frazzon, E.M., Forcellini, F.A., Ivanov, D.: A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. Int. J. Inf. Manag. **49**(3), 86–97 (2019). https://doi.org/10.1016/j.ijinfomgt.2019.03.004
7. Chen, D.L.: Judicial analytics and the great transformation of American Law. Artif. Intell. Law **27**(1), 15–42 (2019). https://doi.org/10.1007/s10506-018-9237-x
8. Choudhury, A.M., Nur, K.: A machine learning approach to identify potential customer based on purchase behavior. In: 1st International Conference on Robotics, Electrical and Signal Processing Techniques, ICREST 2019, pp. 242–247 (2019). https://doi.org/10.1109/ICREST.2019.8644458
9. Christopher, M., Holweg, M.: "Supply chain 2.0": managing supply chains in the era of turbulence. Int. J. Phys. Distrib. Logist. Manag. **41**(1), 63–82 (2011). https://doi.org/10.1108/09600031111101439
10. Diaz-Rozo, J., Bielza, C., Larrañaga, P.: Machine learning-based CPS for clustering high throughput machining cycle conditions. Procedia Manuf. **10**, 997–1008 (2017). https://doi.org/10.1016/j.promfg.2017.07.091
11. Ferreira, K.J., Lee, B.H.A., Simchi-Levi, D.: Analytics for an online retailer: demand forecasting and price optimization. Manuf. Serv. Oper. Manag. **18**(1), 69–88 (2016). https://doi.org/10.1287/msom.2015.0561
12. Gyulai, D., Pfeiffer, A., Nick, G., Gallina, V., Sihn, W., Monostori, L.: Lead time prediction in a flow-shop environment with analytical and machine learning approaches. IFAC-PapersOnLine **51**(11), 1029–1034 (2018). https://doi.org/10.1016/j.ifacol.2018.08.472
13. Halde, R.R., Deshpande, A., Mahajan, A.: Psychology assisted prediction of academic performance using machine learning. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016—Proceedings, pp. 431–435 (2017). https://doi.org/10.1109/RTEICT.2016.7807857
14. Hofmann, E., Rutschmann, E.: Big data analytics and demand forecasting in supply chains: a conceptual analysis. Int. J. Logist. Manag. **29**(2), 739–766 (2018). https://doi.org/10.1108/IJLM-04-2017-0088
15. Jayadi, R., Jayadi, R., Firmantyo, H.M., Dzaka, M.T.J., Suaidy, M.F., Putra, A.M.: Employee performance prediction using Naïve Bayes. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), vol. 8(6), pp. 8–12 (2019)
16. Knoll, D., Prüglmeier, M., Reinhart, G.: Predicting future inbound logistics processes using machine learning. Procedia CIRP **52**, 145–150 (2016). https://doi.org/10.1016/j.procir.2016.07.078
17. Liu, C., Shu, T., Chen, S., Wang, S., Lai, K.K., Gan, L.: An improved grey neural network model for predicting transportation disruptions. Expert Syst. Appl. **45**, 331–340 (2016). https://doi.org/10.1016/j.eswa.2015.09.052
18. Mani, V., Delgado, C., Hazen, B.T., Patel, P.: Mitigating supply chain risk via sustainability using big data analytics: evidence from the manufacturing supply chain. Sustainability (Switzerland) **9**(4) (2017). https://doi.org/10.3390/su9040608
19. Meidan, Y., Lerner, B., Rabinowitz, G., Hassoun, M.: Cycle-time key factor identification and prediction in semiconductor manufacturing using machine learning and data mining. IEEE Trans. Semicond. Manuf. **24**(2), 237–248 (2011). https://doi.org/10.1109/TSM.2011.2118775
20. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. **6**(7), 264–269 (2009)
21. Mortazavi, A., Arshadi Khamseh, A., Azimi, P.: Designing of an intelligent self-adaptive model for supply chain ordering management system. Eng. Appl. Artif. Intell. **37**, 207–220 (2015). https://doi.org/10.1016/j.engappai.2014.09.004

22. Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., Vasilakis, C.: Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. Eur. J. Oper. Res. **290**(1), 99–115 (2021). https://doi.org/10.1016/j.ejor.2020.08.001
23. Ojha, R., Ghadge, A., Tiwari, M.K., Bititci, U.S.: Bayesian network modelling for supply chain risk propagation. Int. J. Prod. Res. **56**(17), 5795–5819 (2018). https://doi.org/10.1080/00207543.2018.1467059
24. Pappalardo, L., Cintia, P.: Quantifying the relation between performance and success in soccer. Adv. Complex Syst. **21**(3–4), 1–30 (2018). https://doi.org/10.1142/S021952591750014X
25. Pereira, M.M., Frazzon, E.M.: A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains. Int. J. Inf. Manag. **57**, 102165 (2021). https://doi.org/10.1016/j.ijinfomgt.2020.102165
26. Priore, P., Ponte, B., Rosillo, R., de la Fuente, D.: Applying machine learning to the dynamic selection of replenishment policies in fast-changing supply chain environments. Int. J. Prod. Res. **57**(11), 3663–3677 (2019). https://doi.org/10.1080/00207543.2018.1552369
27. Sabbeh, S.F.: Machine-learning techniques for customer retention: a comparative study. Int. J. Adv. Comput. Sci. Appl. **9**(2), 273–281 (2018). https://doi.org/10.14569/IJACSA.2018.090238
28. Saradhi, M.V.V., Nageswara Prasad, K.S.: Predictive analytics for obstructive sleep apnea detection. Int. J. Innov. Technol. Explor. Eng. **8**(9 Special Issue 2), 568–573 (2019). https://doi.org/10.35940/ijitee.I1118.0789S219
29. Shang, Y., Dunson, D., Song, J.S.: Exploiting big data in logistics risk assessment via Bayesian nonparametrics. Oper. Res. **65**(6), 1574–1588 (2017). https://doi.org/10.1287/opre.2017.1612
30. Srividya, M., Mohanavalli, S., Bhalaji, N.: Behavioral modeling for mental health using machine learning algorithms. J. Med. Syst. **42**(5) (2018). https://doi.org/10.1007/s10916-018-0934-5
31. Tufano, A., Accorsi, R., Manzini, R.: Machine learning methods to improve the operations of 3PL logistics. Procedia Manuf. **42**(2019), 62–69 (2020). https://doi.org/10.1016/j.promfg.2020.02.023
32. Wang, J., Yang, J., Zhang, J., Wang, X., Zhang, W. (Chris): Big data driven cycle time parallel prediction for production planning in wafer manufacturing. Enterp. Inf. Syst. **12**(6), 714–732 (2018). https://doi.org/10.1080/17517575.2018.1450998
33. Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., Kyek, A.: Optimization of global production scheduling with deep reinforcement learning. Procedia CIRP **72**, 1264–1269 (2018). https://doi.org/10.1016/j.procir.2018.03.212
34. Xia, M., Zhang, Y., Weng, L., Ye, X.: Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. Knowl.-Based Syst. **36**, 253–259 (2012). https://doi.org/10.1016/j.knosys.2012.07.002

Optimizing the Supply of Eucalyptus for Agroindustry: A Mathematical Programming Approach

Moises Knaut Tokarski and José Eduardo Pécora Jr.

Abstract Eucalyptus is the raw material for sectors such as furniture and cellulose, but it is also an important input for thermal energy, being used in several other industrial segments. The high price of agricultural commodities and the low price of eucalyptus affect the renewal of forest areas. The risk of supply in Paraná (Brazil), makes the cereal producers located in the state seek to implement strategies to ensure the supply of eucalyptus chips for drying soybeans and corn. The present work proposes a mathematical model for optimizing the allocation of forest units, reducing costs of leasing, and transporting chips between forest stands and grain receiving units in an 8-year production cycle, considering cuts in forests with ages ranging from 5 to 8 years. The model was validated using a case study of a company located in the Paraná state, in which the optimization guaranteed the supply of chips in 10-grain receiving units and resulted in the choice of 7 out of 14 forest units available for leasing. The model also allows the analysis of the company's business policies, such as verifying the impact of changing the minimum area for harvesting, or minimum transportation lot between a point of origin and a single destination.

Keywords Chipping · Eucalyptus · Allocation · Optimization · Grain receiving

1 Introduction

Eucalyptus has several economic uses, both in the furniture sectors, as raw material in the pulp and paper chain, and as an input in the industrial process of several other activities, such as for drying soybeans and corn. In Paraná, the second-largest grain producer in Brazil, the supply of eucalyptus is threatened in the coming years, because there is no replacement of forest planting in the state. The competition of

M. K. Tokarski · J. E. Pécora Jr. (✉)
Universidade Federal do Paraná, Av. Prefeito Lothário Meissner, 632, Jardim Botânico, Curitiba, PR, Brazil
e-mail: pecora@ufpr.br

M. K. Tokarski
e-mail: moises.knaut@ufpr.br

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_32

areas with agricultural commodities, which break monthly record prices, the low value paid for wood, due to a reflection of the high standing wood stock observed in previous years, added to the intensification of the verticalization of pulp industries in the region, makes that there is risk in the balance between supply and demand for wood in the coming years.

Supply strategies are increasingly essential to avoid shortages or exposure to exorbitant future prices of eucalyptus for any type of use. We can observe in the market, as well as in the literature, a great organization of the pulp and paper industry about the planning and optimization of eucalyptus production processes, both in terms of scheduling and route composition to meet the industry demand, as presented in [1], as well as in the transport planning respecting seasonality restrictions due to snow and ice in Nordic countries, proposed by [2], and also in the optimal allocation of areas for planting, according to [3].

However, other activities that do not have forestry as the main component of the business, such as the marketing of grains, often neglect the strategic management of this input. There are also no studies focusing on process optimization for the supply of eucalyptus wood chips for energy use in grain drying, or in other industrial processes in academia.

We propose an adaptation of the linear programming model presented in [3], to guarantee the optimal allocation of forest production units, reducing costs of leasing and transporting chips between the points of origin and the grain receiving units of a Parana state company that uses eucalyptus chips for drying soybeans and corn.

The research methodology is provided as normative axiomatic modeling, according to [4], as it is an optimization based on a mix of empirical data and data generated for result validation.

The next section describes the proposed problem, while in the third section the mathematical model is presented. In the fourth section, the model is tested and validated in a case study. Finally, the fifth section presents the conclusions of the work.

2 Problem Description

The Paraná State (Brazil southern) has two main agricultural cycle crops in one year, the soybeans on summer and second-crop corn cultivation (planted after the soybean harvest). These grains need to be dried to 13% of humidity to be able to commercialization and preserve the quality conditions.

The study [5] indicate the average technical coefficients for drying soy and corn. According to this report, each crop has an average reference humidity for receiving the product from the crop. For soybean, the reference value is 18%, and for corn 22%. Table 1 presents the consumption of eucalyptus chips, with a humidity of 35%, for the drying of 100 tons of each crop.

Normally, in Brazil southern, the cutting of eucalyptus for energy purposes occurs between 5 and 8 years of age of the forest mass. To achieve an annual harvest to stably

Table 1 Consumption of eucalyptus chips (35% humidity) for soy and corn drying

Culture	Culture reference humidity (%)	Product quantity (t)	Eucalyptus chips consumption (t)
Soybean	18	100	1
Corn	22	100	3.375

Source Adapted from [5]

Table 2 Average wood (m^3) and chip production (m^3 and t) per hectare and age in Paraná

Age (years)	Wood (m^3)	Chips (m^3)	Chips (t)
5	200.0000	587.8571	176.3571
6	240.0000	705.4286	211.6286
7	280.0000	823.0000	246.9000
8	320.0000	940.5714	282.1714

Source Adapted from [6]

meet the drying needs of the grain receiving units, the planting should be staggered in different areas, whether continuous or not. Considering the average annual increment of the state of Paraná of 40 m^3 per hectare per year [6], the amount of wood available in one hectare varies from 200 to 320 m^3, according to the age of the massif. Empirical data collected with companies operating in the state point out that by chopping one cubic meter of eucalyptus it is obtained approximately 2.94 m^3, which is equivalent to 300 kg of chips, with 35% humidity. The eucalyptus and chip production data per hectare according to age are presented in Table 2.

The study [7] shows that the transportation of the chips is one of the main costs associated with the promotion of this raw material in Paraná cooperatives, adding on average R$ 1.84 per kilometer per ton.

Figure 1 demonstrates how the proposed optimization model works when deciding which of the listed areas should be leased and in which years there will have to be a harvest to guarantee the supply of the grain units. In the example illustrated, one can see the choice of forest units 1 and 3, origin 2 would not be leased. It can also be observed that the origins complete the cycle, according to their ages, in supplying the consumer units. Besides the rental value, the model considers the cost per kilometer and ton of chips transported between the points of origin and consumption, reducing the total cost of the operation.

The model enables strategic decision making in the medium term, not addressing the operational part, such as harvest periods, nor the order of loading trucks with chips. For these activities, we suggest adapting the model proposed by [1], which foresees the weekly scheduling of truck loading to meet wood consumption. The system proposed in the current study considers only the rental value and the average cost of displacement, not considering the investment in the acquisition of trucks, nor the marginal cost of chip production operation, i.e., the amount that would be spent

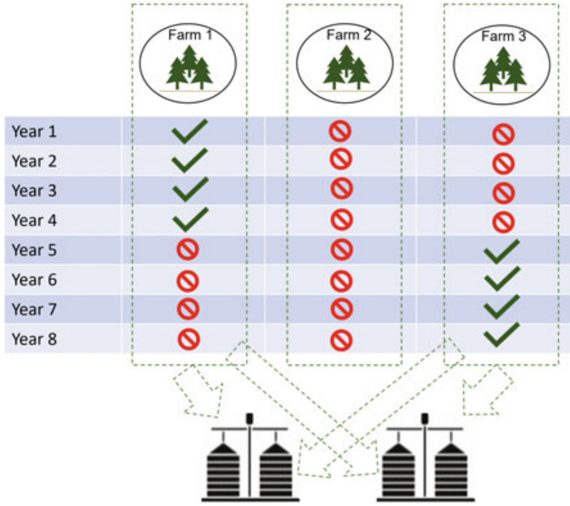


Fig. 1 Decision-making scheme of the proposed model. *Source* Authors (2021)

to produce one ton of chip more than the capacity of the set of equipment responsible for harvesting and chipping the wood.

3 Model

This section presents the mathematical linear programming model proposed to solve this problem, divided into 4 items: sets, parameters, variables, and the model itself.

3.1 Sets

- O (origin points): $o \in O = \{1, \ldots, N\}$
- D (target points): $d \in D = \{1, \ldots, N\}$
- A (years): $a \in A = \{1, \ldots, 8\}$.

3.2 Parameters

- Cost between origin points "o" and chip consumption "d": C_{od}.
- Demand at destination point "d" in the year "a": R_{da}.
- Production per hectare in the year "a" in origin "o": K_{oa}.
- Leasing cost per hectare per year: W_{oa}.
- Total area, in hectares, by origin: H_o.

3.3 Variables

- Quantity of chips between points, expressed continuously: Q_{oda}.
- Allocation of harvest from the leased area in the year, being binary, with value one when area "o" is active in the year "a". Otherwise, the value will be zero: Y_{oa}.
- Allocation of the forest mass leased in the period, being binary, with value one when the area is in the full cycle (eight years). Otherwise, the value will be zero: V_o.
- Harvested area per year "a" in origin "o", continuous variable: N_{oa}.
- If origin "o" will deliver to destination "d" in the year "a", binary variable, being one when active and zero otherwise: B_{oda}.

3.4 Model

$$Min\ Z = \sum_{a \in A} \sum_{d \in D} \sum_{o \in O} C_{od} * Q_{oda} + \sum_{o \in O} W_o * V_o + \sum_{a \in A} \sum_{o \in O} N_{oa} \tag{1}$$

Subject to:

$$\sum_{o \in O} Y_{oa} \leq \sum_{o \in O} V_o, \quad \forall a \tag{2}$$

$$\sum_{d \in D} Q_{oda} \leq K_{oa} * N_{oa}, \quad \forall o, a \tag{3}$$

$$N_{oa} \leq H_o * Y_{oa}, \quad \forall o, a \tag{4}$$

$$Y_{oa} \leq V_o, \quad \forall o, a \tag{5}$$

$$\sum_{a \in A} Y_{oa} \geq V_o, \quad \forall o \tag{6}$$

$$\sum_{a \in A} N_{oa} \leq H_o * V_o, \quad \forall o \tag{7}$$

$$\sum_{o \in O} Q_{oda} \geq R_{da}, \quad \forall d, a \tag{8}$$

$$Q_{oda} \geq B_{oda} * 30, \quad \forall o, d, a \tag{9}$$

$$Q_{oda} \leq B_{oda} * K_{oa} * H_o, \quad \forall o, d, a \tag{10}$$

$$B_{oda} \leq Y_{oa}, \quad \forall o, a, d \tag{11}$$

$$N_{oa} \geq 5 * Y_{oa}, \quad \forall o, a \tag{12}$$

Equation (1) presents the objective function that can be divided into three parts: the first is the sum of displacement costs between the origin and destination points multiplied by the amount of chips transported. The second is the sum of the rental value of each forest unit, when it is allocated ($V_o = 1$). It is considered that the harvesting cost per hectare is the same, regardless of the area in which it is occurring, thus the third part of the equation reduces the harvested area, so that unnecessary harvesting does not occur.

The sets of constraints given by Eqs. (2)–(7) refer to the allocation of forest origin units, harvested area and chip production per year. Equation (2) limits the maximum number of active forests stands (harvested) in a year (Y_{oa}) by the total number of active origins in the full cycle (V_o). Equation (3) limits the amount transported from one origin to all destinations by the maximum that the area can produce in that period. Equation (4) imposes that a massif has part, or all, of its area harvested in a period only if that origin is active in that year. Equations (5) and (6) link the annual allocation variables with period allocation, where a source that is not active in the cycle cannot be active in a single year (Eq. 5) and a massif that is active in the cycle, paying rent, must have at least part of its area harvested in at least one year (Eq. 6). Equation (7) makes sure that the total harvested area in the cycle is not greater than the area of the respective massif.

Equation (8) guarantees meeting the demand for chip consumption by the grain storage units. While Eqs. (9)–(12) are sets of constraints that represent the company's policy aspects: sets (9), (10) and (11) are responsible for guaranteeing a minimum lot for each destination, that is, a forest area should allocate at least 30 tons of chips for each grain unit, to fill the truck's capacity and avoid allocating amounts like one or two tons between the points. The last equation represents the minimum area that the company considers feasible to harvest, in this case, at least five hectares should be harvested per year in an active area ($Y_{oa} = 1$).

4 Case Study

This section will present the data and results of the case study based on a company located in the state of Paraná. The data is partly real and partly created for model validation.

4.1 Data

The case study considered 10-grain receiving units belonging to a company based in the Parana state. Both the locations and the static capacities are real. To calculate the estimated need for chips, the potential receipt of soy and corn grains was used, as well as the need for chips for drying them, described in Table 3 of this work. Table 3 presents the consumption data of each point for the first year of the cycle, with constant consumption throughout the analysis cycle, repeating the values for the other years. However, the model can calculate the variable supply if different demands per year are informed.

The 14 origin points used were randomly generated and their respective areas and ages in the initial year. Consecutively, the production potential of each area was calculated according to the parameters presented in Table 4. Note that when reaching an age of 8 years during the cycle, the following year is considered as age zero, and the count restarts. In this way, a forest mass that is 8 years old in year 1 will be 6 years old in the last period of the cycle. Table 4 presents the origin data, while Table 5 presents the production capacity per year, filtering only the units that are in cutting season per year, that is, between the ages of 5 and 8 years.

The model was programmed in C# language, by the Visual Studio application and solved by the Gurobi Solver 9.1. To calculate the distance, the Haversine formula was used, which results in the distance in a straight line, considering the curvature of the earth, from the latitude and longitude data of the points. For the cost of displacement, the value of R$ 1.84 per kilometer and per loaded ton was attributed.

Table 3 Consumption points and chip demand in tons per year

Destination	City	State	Static Capacity	Latitude	Longitude	Chip Requirement (t/year)
1	Carambeí	PR	1.800	−24.94194	−50.013006	180.68
2	Carambeí	PR	30.900	−24.94414	−50.501289	3101.59
3	Carambeí	PR	2.810	−24.94274	−50.012986	282.05
4	Carambeí	PR	36.540	−24.94355	−50.501275	3667.70
5	Imbaú	PR	1.180	−24.4405	−50.07637	118.44
6	Imbaú	PR	4.610	−24.4407	−50.07644	462.73
7	Imbaú	PR	1.810	−24.4408	−50.07641	181.68
8	Imbaú	PR	9.620	−24.4404	−50.07644	965.61
9	Imbituva	PR	18.610	−25.2188	−50.05836	1867.98
10	Imbituva	PR	3.540	−25.2183	−50.05841	355.33
11	Ponta Grossa	PR	15.340	−25.15525	−50.013997	1539.75
Total			126.760			12,723.54

Source Authors (2021)

Table 4 Data on the forest masses that can be chosen as sources for the chips

Origin	City	State	Area (ha)	Latitude	Longitude	Age at 1st Year	Lease cycle cost (R$/ha)
1	Carambeí	PR	116	−24.9419	−50.3657	4	R$ 8000.00
2	Carambeí	PR	103	−24.9256	−50.9657	4	R$ 10,000.00
3	Carambeí	PR	95	−24.9445	−50.5016	4	R$ 10,400.00
4	Imbaú	PR	129	−24.4488	−50.0886	5	R$ 7920.00
5	Imbituva	PR	108	−25.5454	−50.0555	8	R$ 8800.00
6	Imbituva	PR	147	−25.5858	−50.0625	8	R$ 8000.00
7	Ponta Grossa	PR	52	−25.1818	−50.0155	1	R$ 8000.00
8	Ponta Grossa	PR	124	−25.1557	−50.0366	1	R$ 8000.00
9	Ponta Grossa	PR	67	−25.149	−50.0145	1	R$ 9600.00
10	Ponta Grossa	PR	84	−25.1545	−50.5152	1	R$ 8880.00
11	Teixeira Soares	PR	111	−25.4188	−50.0566	1	R$ 7600.00
12	Tibagi	PR	79	−24.5157	−50.4586	7	R$ 8000.00
13	Tibagi	PR	48	−24.6878	−50.5859	7	R$ 8000.00
14	Tibagi	PR	115	−24.5987	−50.1515	7	R$ 9240.00

Source Authors (2021)

The base company of this study has well-established policies regarding minimum harvest area and minimum transport lot:

- Minimum area for harvesting: You must only activate an origin if the harvest is equal to or greater than 5 ha.
- Minimum transport lot: There should only be a flow of chips between an origin and a destination if the amount transported is equal to or greater than 30 tons.

The model should reduce the costs of displacement and land leasing, guaranteeing that demand is met and respecting the physical and political restrictions of the process. As a premise, it is considered that the grain receiving units can store all the chips they need in a year, that is, they can receive all the raw material that will be used at any time of the year.

Adaptations to cater for consumption points that do not obey this assumption can be made by adding a subset "P" of destinations with this characteristic in the model, that belong to the set "D" (destination points). In addition, it is also necessary to add a restriction like Eq. (9), changing the multiplier B_{oda} by Y_{oa}. In this way, all active origin in that year should deliver chips to points "P", as presented in Eq. (13).

$$Q_{opa} \geq Y_{oa} * 30, \qquad \forall p, o, a \tag{13}$$

Table 5 Chip's production in tons per hectare per year, according to forest age

Year	Origin	Age	Production (t/ha)	Year	Origin	Age	Production (t/ha)
1	4	5	176.36	5	9	5	176.36
1	5	8	282.17	5	10	5	176.36
1	6	8	282.17	5	11	5	176.36
1	12	7	246.90	6	7	6	211.63
1	13	7	246.90	6	8	6	211.63
1	14	7	246.90	6	9	6	211.63
2	1	5	176.36	6	10	6	211.63
2	2	5	176.36	6	11	6	211.63
2	3	5	176.36	7	5	5	176.36
2	4	6	211.63	7	6	5	176.36
2	12	8	282.17	7	7	7	246.90
2	13	8	282.17	7	8	7	246.90
2	14	8	282.17	7	9	7	246.90
3	1	6	211.63	7	10	7	246.90
3	2	6	211.63	7	11	7	246.90
3	3	6	211.63	8	5	6	211.63
3	4	7	246.90	8	6	6	211.63
4	1	7	246.90	8	7	8	282.17
4	2	7	246.90	8	8	8	282.17
4	3	7	246.90	8	9	8	282.17
4	4	8	282.17	8	10	8	282.17
5	1	8	282.17	8	11	8	282.17
5	2	8	282.17	8	12	5	176.36
5	3	8	282.17	8	13	5	176.36
5	7	5	176.36	8	14	5	176.36
5	8	5	176.36				

Source Authors (2021)

4.2 Results

All model constraints were satisfied, meeting the demand of the grain receiving units, activating 7 of the 14 available forest blocks. The area harvest limits, and their production capacities were respected according to their respective ages. No stock was generated at the points of origin, that is, all the wood harvested was chopped and forwarded to the destination units. The total cost in the period was R$ 9.06 million, of which R$ 4.35 million were for chip transportation and R$ 4.71 million for leasing the areas.

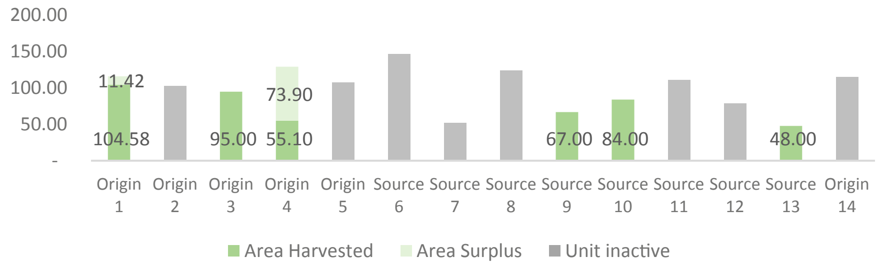


Fig. 2 Forest units with harvested, surplus, and inactive areas. *Source* Authors (2021)

Figure 2 shows the choice of forest production units, as well as their harvested area. It can be observed that units 1 and 4 have a surplus of the unharvested area in the period. In this scenario, this surplus is a waste since the rent paid corresponds to the total area, which is only partially being used. On the other hand, this is an opportunity for the company, which, based on this information, can decide between several options, such as:

- Reviewing the lease contract, adjusting the actual area that will be used.
- Use of the surplus area for other purposes, such as management for furniture.
- Utilization of idle harvesting and chipping equipment (53% of the capacity used per year) and sale of chips to third parties.

Policy analysis. By altering the sets of restrictions that define the company's policies, it is possible to verify the impact of this change on the total cost of operation, serving as a basis for confirming or revising the policies adopted. In this work, the projection of costs will be made with the alteration of the minimum area for harvesting.

Changing the minimum harvest area. Analysis of the impact of the minimum harvest area on cost was done with the change of the minimum area, in constraint Eq. (12). The other constraints were maintained and met. The tested instances and the costs resultants were:

- Instance 0: Original problem, with a limit of 5 ha—cost: R$ 4.36 million.
- Instance 1: 1 ha—cost: R$ 4.35 million.
- Instance 2: 2.5 ha—cost: R$ 4.35 million.
- Instance 3: 7.5 ha—cost: R$ 4.36 million.
- Instance 4: 10 ha—cost: R$ 4.34 million.
- Instance 5: 15 ha—cost: R$ 4.37 million.

As shown on the list, instances 1 and 2 showed no difference about instance 0. One notices that the lowest value found was in instance 4, with a minimum of 10 ha harvested to activate the area. The cost reduction was R$ 5232.42 under these conditions. The units chosen for harvesting were not altered with the modifications and the rental cost remained the same. The main differences observed were an increase in the area harvested at origin 1 and a reduction at origin 4, as shown in Fig. 3. A

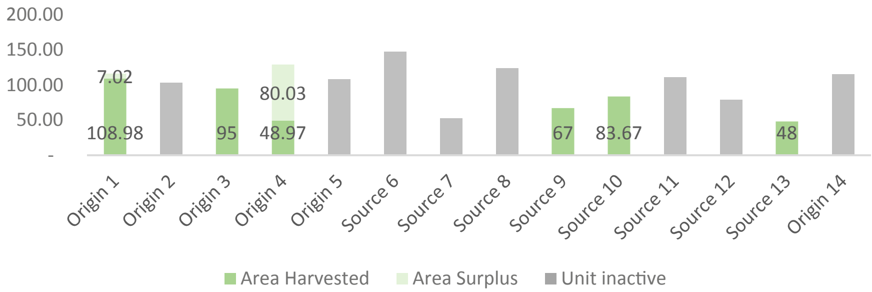


Fig. 3 Forest units with harvested, surplus, and inactive areas—instance 4. *Source* Authors (2021)

secondary gain is the reduction of active forest units in year 4 of the cycle, which in instance 0 were three and in instance 4 were two, reducing the impact of equipment displacement between forest blocks.

5 Conclusion

The proposed model meets the objective of creating a tool that enables decision-making regarding which forest units to lease, what is the ideal distribution of chips about the quantity transported between the origins and the consumption points, reducing the global cost of the operation. Besides this, it also allows the construction of scenarios that make it possible to evaluate the internal policies of the companies, validating them or suggesting changes to further reduce the strategic operational costs of this important activity of biomass promotion for grain drying.

Complementary future work can be carried out on two fronts: the first is the adaptation to consider costs not accounted for in the present work, such as the displacement of harvesting and chipping equipment, or even investment feasibility analysis. The second is the transformation of the model for strategic decisions for tactical operational optimization, as presented by [1], evaluation of their work. Papers outside the formatting will be removed from the evaluation process.

References

1. Bordón M, Montagna J, Corsano G (2020) Mixed-integer linear programming approaches for solving the raw material allocation, routing, and scheduling problems in the forest industry. Int J Ind Eng Comput
2. Sfeir TA, Pecora JE, Ruiz A, Lebel L (2021) Integrating natural wood drying and seasonal trucks' workload restrictions into forestry transportation planning. Omega Int J Manag Sci
3. Rönnqvist M (2003) Optimization in forestry. Math Program Ser B 97:267–284
4. Cauchick-Miguel PA, Fleury A, Mello CHP, Nakano DN, Lima EP, Turrioni JB, Ho LL, Morabito R, Costa SEG, Martins RA, Sousa R, Pureza V (2018) Research methodology in production

engineering and operations management, 3rd edn. In: Coordination of Cauchick-Miguel PA (ed). Elsevier, Rio de Janeiro, Portuguese
5. Custos de Recepção, secagem e limpeza de grãos nº 457 (2020) Cooperative Union and Organization of the State of Paraná—Ocepar, Portuguese
6. Sector study—APRE (2018) Paranaense Association of Forest-Based Companies, Portuguese
7. Strategic dialogue—wood supply in the cooperatives of Paraná (2020) Cooperative Union and Organization of the State of Paraná—Ocepar and Innovatech Consultant, Portuguese

Green Television Supply Chain Under Capital Constraint for Achieving Environmental Sustainability

Song-Man Wu, Felix T. S. Chan, and S. H. Chung

Abstract In recent years, the environment has become a hot topic among the government, society, and enterprises due to global warming, ozone depletion, and air pollution. Businesses are believed to be at the root of most environmental problems. To reduce environmental pollution and promote sustainable development, more and more TV manufacturers are adopting a green supply chain to achieve sustainable development and produce more environmentally friendly television (TV) sets. However, some green TV manufacturers are underfunded in the development and production process. In this paper, a green TV supply chain system composed of a financially constrained manufacturer and a reputable retailer is proposed to help the TV manufacturer successfully produce more green TVs. The supply chain model is established by using two financing tools. A time-dependent residual value is considered in the model, and the later the clearance time, the lower the residual value. We investigate the effects of time-dependent residual value on operational and financing decisions and measure the supply chain system's profit risks. The following results are concluded: (1) The clearance time of unsold TV sets affects the retailer's and the manufacturer's decisions. (2) The financing equilibrium can be obtained under certain conditions. (3) The profit risks of the retailer increase with the order quantity and the clearance time. Numerical analysis is used to verify the derived results.

Keywords Green supply chain management · Environmental sustainability · Capital constraint · Supply chain finance · Time-dependent residual value

S.-M. Wu · F. T. S. Chan (✉) · S. H. Chung
The Hong Kong Polytechnic University, Hung Hum, Hong Kong
e-mail: f.chan@polyu.edu.hk

S.-M. Wu
e-mail: songmanwu@foxmail.com

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_33

1 Introduction

Along with the advance of electronic technology, through the tireless efforts of countless scientists, engineers, the television (TV) from the original large samples of laboratory, image fuzzy, to develop into an entered innumerable family, the main household appliances products, now the ultrathin, image is clear, powerful display products. In addition, due to the seriousness of the problems of pollution and energy consumption, the market share of energy-saving green TV is growing. At the same time, as TVs are constantly updated, more and more out-of-season products fall out of favor and can't be sold during the selling season, so retailers face the problem of excess inventory. Therefore, considering the residual values of out-of-season TV sets has become a problem worth studying of supply chain management in recent years. It can not only reduce the profit loss of retailers but also contribute to environmental protection and sustainable development.

Further, due to the large investment in TV research and development and the high production cost, some TV manufacturers, especially small and medium-sized enterprises (SMEs), are often restricted by the lack of funds in the production and supply process [1]. When these manufacturers run short of capital, downstream buyers are exposed to a shortage of goods, which affects the normal operation of the business. However, due to the lack of collateral or credit, these small manufacturers become more difficult and expensive in the financing process [2]. For solving this problem of capital shortage, supply chain finance (SCF) is widely adopted in the world, especially in developing countries. It uses the loan resources of core enterprises to help the enterprises of the supply chain to obtain financial help. Therefore, how to establish effective financing channels to solve the problem of capital constrains of SMEs has become a major problem. Most of the studies use post-shipment financing, the research on pre-shipment financing is still few. In addition, in many studies on SCF, many researchers ignore the residual value in short-term financing to simplify the calculation. Therefore, in their study, the residual value of unsold products was set to zero. There has been little systematic research in the SCF literature on the impact of residual value, especially based on time-dependent residual value.

Therefore, to fill the above gaps, a green TV supply chain system composed of a financially constrained manufacturer and a reputable retailer is proposed to help the TV manufacturer successfully produce more green TVs. A time-dependent residual value is considered in the model, and the later the clearance time, the lower the residual value. We focus on two financing instruments, namely buyer-supported purchase order financing (BPOF) and advance payment discounting (APD) [3], to ease the capital constraints of the upstream manufacturer. Therefore, our research explores how to manage the corporate financial risks with time-dependent residual values in the green supply chain. Our main contributions are in three aspects: (1) The clearance time of unsold TV sets affects the retailer's and the manufacturer's decisions. (2) The financing equilibrium is BPOF under certain conditions. (3) The profit risks of the retailer increase with the order quantity and the clearance time.

The rest of this research begins with the literature review in Sect. 2. Two models are built in Sects. 3 and 4. Section 5 analyzes the financing equilibrium and profit risk respectively. Numerical experiments are carried out in Sects. 6 and 7 summarizes and concludes the paper.

2 Literature Review

Two literature streams are relevant to the research: the study on dealing with the enterprise's budget constraint by SCF and the combination of green supply chain management (GSCM) and SCF.

Firstly, SCF is used to deal with the problem of enterprise budget constraints. Recently, Alan and Gaur [4] explored the role of inventory in both bank financing and trade credit. Khan et al. [5] investigated the impact of early-payment financing on supply decisions. Yan et al. [6] studied two financing methods that the loss-averse retailers could provide financial services (loans and investments) to help the cash-strapped supplier and compared the two financing methods. Zhen et al. [7] established an SCF model in which a cash-constrained manufacturer could choose to borrow from a third-party platform, a retailer, or a bank, and studied the manufacturers' financing strategy choice. The research on SCF mainly focuses on the capital constraints of retailers, but less on the capital constraints of upstream enterprises.

Secondly, the closest thing to our work is the combination of GSCM and SCF research. It is widely known that GSCM aims to encourage supply chain parties to consider energy conservation and emission reduction when making decisions. Li et al. [8] established Stackelberg game models to study equilibrium decisions and profit coordination methods of GSCM. Ghosh et al. [9] focused on GSCM and developed a strategy based on advanced payment policies and trade credit facilities. Luo et al. [10] took into account both supplier financing and bank financing to handle the capital constraints. Xu and Fang [11] developed an emission-dependent model with partial credit guarantee and trade credit. Abdel Basset et al. [12] and Lin and He [13] also integrated SCF with GSCM.

3 Model Description

This study considers a newsvendor framework that there are three key participants: an SME manufacturer, a retailer, and a financial institution, all of whom are risk neutral. The manufacturer is strapped for cash and these parties in this supply chain collaborate to deal with the financial problem by using two supply chain financing tools. APD is that the retailer can finance the manufacturer by pre-paying for products at a discount before delivery. BPOF means that the manufacturer obtains funds from a financial institution based on a purchase order with the established retailer's guarantee. Figure 1 displays a sequence of events. The notations are in Table 1.

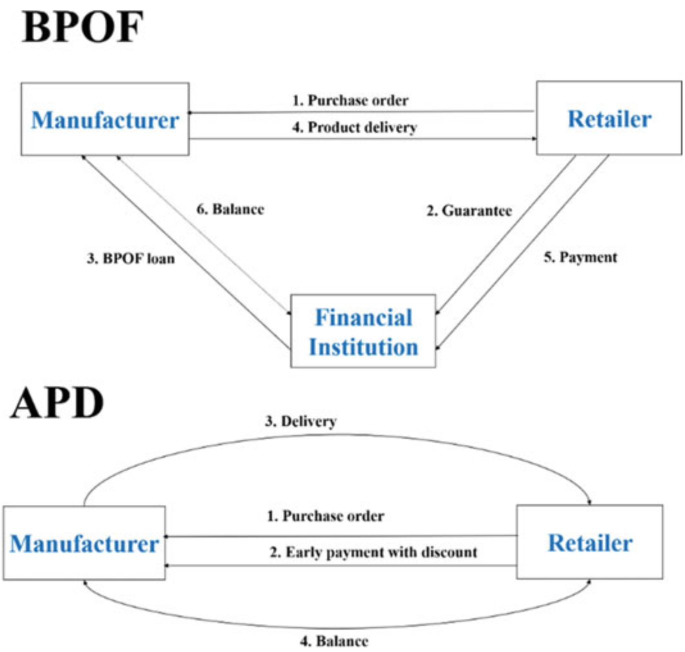


Fig. 1 The sequence of events of two financing instruments

Table 1 Summary of notations

Notation	Description
π	Expected profit
A	Asset level
L	Short-term debt
p	Selling price
α	Proportional distress cost
γ	Proportional liquidation cost
δ	A portion of financial institution's loss
cp	Production cost
ck	Capacity cost
r	Interest rate
s	Time-dependent residual value
K	Capacity level
q	Order quantity
w	Wholesale price
d	Discount rate
λ	A portion of purchase order value

After the sales season ends, any leftover inventory is salvaged at s and it is time varying. The manufacturer has internal assets As and short-term debt Ls [14]. Ls is certain and it will expire before the sales peak. The capital market in our study is imperfect. That is, when a firm fails to pay its debts, it can either be liquidated or negotiated with creditors in an expensive restructuring process. In the case of liquidation, the cost of financial default is a proportion $1 - \gamma$ $(0 < \gamma < 1)$ of firm value. In the case of reorganization, the cost of financial distress is a proportion $1 - \alpha$ $(0 < \alpha < 1)$ of raised capital [3].

4 Models

4.1 APD Model

By adopting APD, the manufacturer's profit and capacity in APD is

$$\pi_m^{apd} = \left[w(1-d) - c_p\right]N - c_k K - (1-\alpha)[L_s - A_s + c_k K - w(1-d)q]^+ \tag{1}$$

where $N = \min(q, K)$. When $L_s - A_s + c_k K - w(1-d)q < 0$, the manufacturer continues to operate. Conversely, if $L_s - A_s + c_k K - w(1-d)q > 0$, the manufacturer goes to reorganization.

The retailer's expected profit is:

$$\begin{aligned}\pi_r^{apd} &= p\mathrm{E}\min[D, N] - w(1-d)N + (a - bt)[N - D]^+ \\ &\quad - (1-\alpha)[L_r - A_r + w(1-d)q]^+\end{aligned} \tag{2}$$

Under APD, q^{apd*} satisfies $p - (p - a + bt)F\left(q^{apd*}\right) = w(1-d)$ in continuation and in the reorganization the optimal order quantity satisfies $p - (p - a + bt)F\left(q^{apd*}\right) = w(1-d)(2-\alpha)$. The retailer balances the marginal effect of APD against the unit financial distress cost.

Proposition 1 *(1) Under APD, the earlier the clearance time of the TV, the buyer will order more products, which will bring higher profits; (2) Clearance time is inversely proportional to the discount rate.*

Proposition 1 first indicates that the clearance time of the TVs negatively affects the retailer's order amount and profit. This is because that the earlier the clearance, the higher the residual value, the less loss to the retailer's profit. Conversely, goods with a low residual value that are cleared late are ordered in fewer quantities. In addition, the lower the residual value of the TVs, the lower the discount rate offered by the manufacturer. Because when the residual value of the TV becomes lower, the financial risk of the retailer becomes higher. Therefore, the seller should offer a small d to the buyer to keep higher purchasing prices and thus release financial risks.

4.2 BPOF Model

In this case, the manufacturer's profit is

$$\pi_m^{bpof} = (w - c_p)N - c_k K - \lambda wqr - (1-\alpha)(L_s - A_s + c_k K - \lambda wq)^+ \quad (3)$$

Here $N = \min(q, K)$. When $L_s - A_s + c_k K - \lambda wq < 0$, the manufacturer continues to operate. Conversely, if $L_s - A_s + c_k K - \lambda wq > 0$, the manufacturer goes to reorganization.

The buyer's expected profit function is:

$$\pi_r^{bpof} = pE\min[D, N] - wN + (a - bt)[N - D]^+ - \int_{\underline{A}_S}^{\tilde{A}_S} \delta\left\{\lambda wq - \gamma\begin{bmatrix}(w - c_p)N - c_k K \\ -\lambda wqr + A_s - L_s\end{bmatrix}\right\}\phi(A_s)dA_s \quad (4)$$

And the optimal order quantity is $q^{bpof*} = F^{-1}\left(\frac{p - w - (\delta\lambda w - \delta\gamma w + \delta\gamma c_p + \delta\gamma\lambda wr)\left[\Phi(\tilde{A}_S) - \Phi(\underline{A}_S)\right]}{p - a + bt}\right)$.

Proposition 2 *In BPOF, both π_r^{bpof} and q^{bpof*} are related to the time-dependent residual value.*

Proposition 2 suggests that if the order quantity exceeds the market demand, that is, there is surplus inventory. Items with high residual values which are cleared early will encourage to order more products because high residual values can help reduce losses and maximize profits.

5 Financing Equilibrium and Profit Risk Analysis

If either of the two financing instruments can satisfy the upstream manufacturer's financial requirements, the downstream buyer will select one of them in the single financing condition.

Theorem 1 *In single financing, there is a unique threshold of the retailer's internal asset level ω_r exists, which makes the buyer preferred to choose BPOF if and only if $A_r < \omega_r$. Otherwise, APD is chosen. Besides, the value ω_r is in inverse proportion to t.*

Theorem 1 reveals that ω_r with early clearance time is higher than ω_r with late clearance time. The reason is that goods with high residual value will encourage the retailer to have more orders, and at the same time, the increase of the order quantity will increase the inventory risks of the retailer, making the inventory in the backlog state. Therefore, if the retailer buys goods with a high residual value, the buyer's asset level should be higher, so enough funds can be provided by the retailer to the manufacturer via APD financing.

The profit risks if the retailer can be analyzed via a mean–variance analysis [15]. The variance of the profit in APD is:

$$V\left(\pi_r^{apd}\right) = (p - a + bt)^2\left[2Nn - 2\int_0^N xF(x)dx - n^2\right] \tag{5}$$

where $N = \min(q, K), n = \int_0^N F(x)dx$.

In addition, the variance of the profit in BPOF is:

$$V\left(\pi_r^{bpof}\right) = (p - a + bt)^2\left[2Nn - 2\int_0^N xF(x)dx - n^2\right] \tag{6}$$

where $N = \min(q, K), n = \int_0^N F(x)dx$.

Proposition 3 *$V(\pi_r^{apd})$ and $V(\pi_r^{bpof})$ are increasing with q and t.*

Proposition 3 illustrates the retailer's profit risks under two financing tools increases with the order quantity and the clearance time, that is, the higher order quantity leads to higher profit risks and the longer TV sets are cleared, the greater the risk to the retailer's profits.

6 Numerical Analysis

In this section, the effects of time-dependent residual values on the manufacturer's discount rate are investigated through some computational experiments. $p = \$580$, $\alpha = 0.85$, $ck = cp = \$120$, $a = 300$ and $b = 10$. D obeys a normal distribution N(1000, 100). Further, in our paper, t is set to vary between 0 and 30.

We show the effect of time-dependent residual values on the manufacturer's discount rate in Fig. 2. As can be seen from Fig. 2, as the clearance increasing, the discount rate decreases. That is to say, when there is unsold TV, if t is large (i.e., $t = 27$), the discount rate at this time is less than the discount rate at $t = 6$. Since the residual value of the TV is low, the retailer will endure more losses during the clearance period. Hence, facing with high financial risks of the buyer, the seller should offer a lower discount rate to maintain a higher wholesale price to reduce his financial risks.

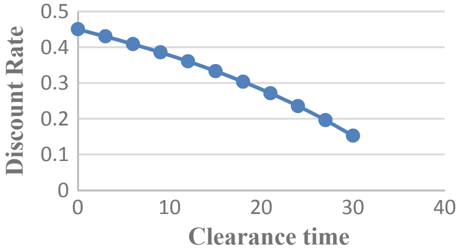


Fig. 2 Impact of t on supplier's discount rate

7 Conclusion

To help the cash-strapped supplier obtain financing for producing more green TV sets, we studied an SCF systeam consisting of a manufacturer, a retailer, and a financial institution, and analyzed the effect of the time-dependent residual value on the participants' optimal decision-making and financing strategy. Some important results and management implications are obtained: (1) The residual value of TV sets varies with the clearance time, which affects the retailer's optimal order quantity as well as the manufacturer's discount rate; (2) When the clearance time is lower than a certain threshold or the retailer's internal asset level is lower than a certain threshold, the financing equilibrium is BPOF; (3) The profit risk of the retailer increases with the increase of order quantity and the clearance time. Future research directions can explore the risk preference and information asymmetry of supply chain participants in the model.

References

1. Vandenberg, P.: Adapting to the financial landscape: evidence from small firms in Nairobi. World Dev. **31**(11), 1829–1843 (2003)
2. Kouvelis, P., Zhao, W.: Financing the newsvendor: supplier vs. bank, and the structure of optimal trade credit contracts. Oper. Res. **60**(3), 566–580 (2012). https://doi.org/10.1287/opre.1120.1040
3. Zhao, L., Huchzermeier, A.: Managing supplier financial distress with advance payment discount and purchase order financing. Omega (2018). https://doi.org/10.1016/j.omega.2018.10.019
4. Alan, Y., Gaur, V.: Operational investment and capital structure under asset-based lending. Manuf. Serv. Oper. Manag. **20**(4), 637–654 (2018). https://doi.org/10.1287/msom.2017.0670
5. Khan, M.A.A., Shaikh, A.A., Panda, G.C., Konstantaras, I., Cárdenas-Barrón, L.E.: The effect of advance payment with discount facility on supply decisions of deteriorating products whose demand is both price and stock dependent. Int. Trans. Oper. Res. **27**(3), 1343–1367 (2019). https://doi.org/10.1111/itor.12733
6. Yan, N., Jin, X., Zhong, H., Xu, X.: Loss-averse retailers' financial offerings to capital-constrained suppliers: loan vs. investment. Int. J. Prod. Econ. **227** (2020). https://doi.org/10.1016/j.ijpe.2020.107665

7. Zhen, X., Shi, D., Li, Y., Zhang, C.: Manufacturer's financing strategy in a dual-channel supply chain: third-party platform, bank, and retailer credit financing. Transp. Res. Part E Logist. Transp. Rev. **133** (2020). https://doi.org/10.1016/j.tre.2019.101820
8. Li, P., Rao, C., Goh, M., Yang, Z.: Pricing strategies and profit coordination under a double echelon green supply chain. J. Clean. Prod. **278**, 123694 (2021). https://doi.org/10.1016/j.jclepro.2020.123694
9. Ghosh, P.K., Manna, A.K., Dey, J.K., Kar, S.: Supply chain coordination model for green product with different payment strategies: a game theoretic approach. J. Clean. Prod. **290** (2021). https://doi.org/10.1016/j.jclepro.2020.125734
10. Luo, Y., Wei, Q., Ling, Q., Huo, B.: Optimal decision in a green supply chain: bank financing or supplier financing. J. Clean. Prod. **271**, 122090 (2020). https://doi.org/10.1016/j.jclepro.2020.122090
11. Xu, S., Fang, L.: Partial credit guarantee and trade credit in an emission-dependent supply chain with capital constraint. Transp. Res. Part E Logist. Transp. Rev. **135**, 101859 (2020)
12. Abdel Basset, M., Mohamed, R., Sallam, K., Elhoseny, M.: A novel decision-making model for sustainable supply chain finance under uncertainty environment. J. Clean. Prod. **269**, 122324 (2020). https://doi.org/10.1016/j.jclepro.2020.122324
13. Lin, Q., He, J.: Supply chain contract design considering the supplier's asset structure and capital constraints. Comput. Ind. Eng. **137**, 106044 (2019)
14. Babich, V.: Independence of capacity ordering and financial subsidies to risky suppliers. Manuf. Serv. Oper. Manag. **12**(4), 583–607 (2010). https://doi.org/10.1287/msom.1090.0284
15. Choi, T.M., Li, D., Yan, H.: Mean–variance analysis of a single supplier and retailer supply chain under a returns policy. Eur. J. Oper. Res. **184**(1), 356–376 (2008)

Fuzzy Multi-objective Mathematical Modelling for Distribution Planning

Peter Adjei, Anibal Careaga-Campos, Gursel Suer, and Omar Alhawari

Abstract This paper discusses a distribution planning model with *n* distribution centers (DCs) and *m* customers. Ideally, each DC serves customers in its zone. Each customer order may include multiple items. However, as the need arises, customers can also be served from other DCs to avoid delays to the customers. Besides, items in a customer's order may also come from multiple DCs not to violate safety stock values established for each item in each warehouse. Furthermore, there are alternative transportation modes available with different costs and CO_2 emission. Environmental issues have become increasingly popular and various studies have been proposed in the literature in recent years. Many corporates are increasing their sustainability efforts and becoming more responsible members of the society. We assume stochastic transportation leadtimes. They are assumed to be normally distributed with known mean and standard deviation values. There are multiple objectives considered in this study and they are: (1) minimize total shipping cost, (2) minimize probability of tardiness, (3) minimize the environmental impact of transportation, (4) minimize safety stock violations. A multi-objective fuzzy mixed-integer-non-linear programming model is developed to handle all these objectives such that satisfaction level is maximized. The task is to determine which DC(s) will serve each customer. Two experiments are conducted, (i) the same network is experimented with multiple periods where inventory levels of items are revised at the end of each period based on the shipment made, (ii) multiple experiments are conducted with various parameters including number of customers, number of items in orders, transportation leadtime distributions, starting inventory levels, safety stock values and fuzzy operators. One of the important features of this study is that it combines both stochastic and fuzzy parameters in the same mathematical model, which has been rarely found in the literature.

P. Adjei · A. Careaga-Campos · G. Suer · O. Alhawari (✉)
Department of Industrial and Systems Engineering, Russ College of Engineering and Technology, Ohio University, Athens, OH 45701, USA
e-mail: alhawari@ohio.edu

G. Suer
e-mail: suer@ohio.edu

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_34

Keywords Fuzzy integer programming · Supply chain network design · Sustainability

1 Introduction

A supply chain includes primary internals elements such as sourcing, production, distribution, etc. [12]. As a focus of this paper, the distribution term refers to moving and storing items from the point of origin (suppliers) to the point of consumption (customers) in a supply chain [5]. This assures the importance of distribution planning in how products are retrieved and delivered from distribution centers to customers, which in turns impacts supply chain performance and customer satisfaction levels. Benjamin [2] addressed the issue of distribution planning in a supply chain by selecting the transportation mode considering multiple suppliers, customers, and a single product. Later, Zhou et al. [19], used genetic algorithm to tackle the problem of allocating customers to multiple distribution centers in a supply chain. Chan and Chung [3] proposed a genetic algorithm to tackle the problem of distributing products from number of sources to demand points. Farahani and Elahipanah [9] discussed the issue of just-in-time distribution in which a multi-objective model was proposed to achieve three objectives, and they are minimizing costs, backorders, and surpluses of products in a multi-period environment.

In recent years, Tang et al. [18] studied the issue of distribution planning by considering multi-layer distribution system between suppliers and customers. In their work, a mixed integer linear model was used to solve the optimization problem of determining the optimal distribution centers and then assign customers to such facilities. Hsiao et al. [10] studied the distribution planning problem is a food supply chain with focus on identifying the best plan that ensures meeting customer requirements. A heuristic approach was developed to achieve objectives of minimizing the distribution costs. Mohammed and Wang [14] addressed the distribution planning issue in supply chains in which various objectives such as minimizing the transportation costs, amount of CO_2 emissions and product distribution time. Consequently, a fuzzy mathematical model was proposed to achieve the satisfactory level of the objectives considered. Molla-Alizadeh-Zavardehi and Shoja [15] stated that distribution performance in supply chains is related to direct or indirect shipment. The direct shipment is considered when products are distributed from origin point to customers while the indirect shipment is not. Further, the authors mentioned that the classic transportation issues consider factors such as times and costs that are usually certain. Therefore, most of the existing optimization models ignore the uncertainty matter. According to Molla-Alizadeh-Zavardehi and Shoja [15], uncertain environment is resulted from different considerations such as traffic jam, machine breakdown, materials quality issues, etc. Based on that, the authors developed a fuzzy mathematical model for a responsive supply chain including multi products and direct shipment considering cost coefficients as trapezoidal fuzzy numbers. Cosma et al. [7] addressed the distribution planning in which the transportation problem is considered. The problem was modelled as a distribution network including manufacturers, distribution centers and customers with an objective of minimizing the total distribution costs. Additionally, a heuristic algorithm, as a solution method, was developed to solve such a problem. Azizi and Hu [1] studied the distribution issue in supply chains in which the problems of the location of distribution centers, vehicle routing and direct shipment are considered simultaneously. Thus, a mixed integer linear programming model was developed to tackle the problems considered with an objective of minimizing total costs incurred.

Further, the distrubution planning in supply chains has been addressed in literature including planning level (i.e. operational), transportation mode (i.e. truck), demand nature (i.e. stochastic) and periods of planning (i.e. multi-period) [3, 4, 13, 17]. However, the strategic planning level and deterministic demand were considered by Contesses et al. [6]. Also, Humair et al. [11] considered stochastic lead times and network inventory optimization in which closed-form equations were presented for the safety stock levels in a supply chain. Diabat et al. [8] incorporated uncertain lead times and demand in a problem concerned with determining location of distribution centers, assigning retailers to distribution centers, and minimizing shipment costs accordingly. In their work, Puga et al. [16] studied stochastic lead times and safety stock placement decisions. It was shown that in the case of having short lead times between distributions centers and retailers, distribution centers are more likely to hold safety stock while retailers offer fast delivery. On the other hand, only the retailers hold safety stock considering longer lead times.

As compared to the aforementioned existing literature findings, this paper is distinguished by presenting the operational level of distribution planning in making related decisions such as minimizing shipping cost, probability of tardiness, the environmental impact of transportation and safety stock violations More importantly, the main contribution of this paper is to fill the gap, existed in literature, in which none of the studies combined both stochastic and fuzzy parameters in the same mathematical model. The remainder of the paper is organized as follows. Section 2 describes the problem under study and the importance to why it should be studied. In Sect. 3, the proposed methodology is described. Numerical experiments are presented in Sect. 4. In Sect. 5, we conclude the paper.

2 Problem Statement

The objective of this study is to integrate supply chain modeling for a distribution system contemplating not only the minimization of distance traveled and total cost but including environmental concerns as well as inventory control. The problem explores a distribution system comprised of two distribution centers (DC's) and ten customers (C's). both DC's can deliver to all customers using two different transportation modes. Each transportation mode includes different distances, different lead times, and different CO_2 emissions. Figure 1 shows the distribution network for one of the explored problems.

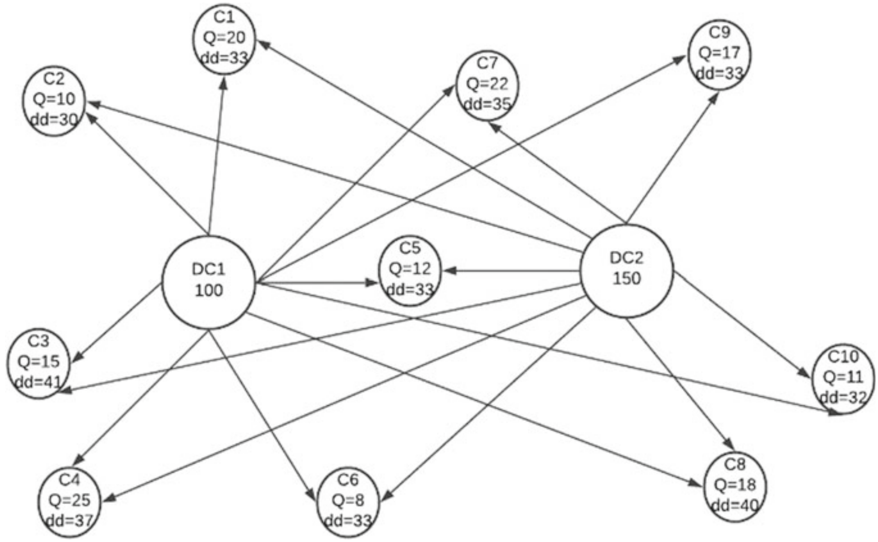


Fig. 1 Distribution network

3 Proposed Methodology

A multi-objective mixed-integer-non-linear programming model is proposed to minimize the distance traveled, minimize the CO_2 emissions associated with selecting alternative transportation mode, minimize the probability of tardiness, and maintain healthy levels of inventory. Transportation time from the manufacturing plant to the customer is stochastic following normal distribution, and for that, it is important for customers to receive the ordered products within the established due date. Therefore, probability of earliness is defined to measure this aspect. In choosing the transportation mode, it is also important to minimize the carbon emission. For instance, truck is the fastest available mode and will increase the probability of earliness; however, it is expensive, and the carbon emission associated with this mode is high.

3.1 Notation

Indices:

j Customer index
m Distribution center index
k Transportation mode index.

Parameters:

NC Number of customers

NDC	Number of distribution centers
NK	Number of transportation modes
Lb1	Lower bound for satisfaction level of performance measure 1
Ub1	Upper bound for satisfaction level of performance measure 1
Lb2	Lower bound for satisfaction level of performance measure 2
Ub2	Upper bound for satisfaction level of performance measure 2
Lb3	Lower bound for satisfaction level of performance measure 3
Ub3	Upper bound for satisfaction level of performance measure 3
Lb4	Lower bound for satisfaction level of performance measure 4
Ub4	Upper bound for satisfaction level of performance measure 4
Dj	Due date for customer *j*
Qj	Quantity ordered from customer *j*
CEmjk	Carbon emission associated with delivery from distribution center m to customer j by using transportation mode k
tmjk	Transportation cost from distribution center *m* to customer *j* by using transportation mode *k*
μmjk	Mean of transportation Leadtime from distribution center *m* to customer *j* when transportation mode *k* is used
σmjk	Standard deviation of transportation Leadtime from distribution center *m* to customer *j* when transportation mode *k* is used
Im,o	Current inventory level in distribution center *m*.

Decision Variables:

λ_1	Satisfaction level—cost
λ_2	Satisfaction level—probability of tardiness
λ_3	Satisfaction level—carbon emission
λ_4	Satisfaction level—inventory level
Xmjk	1 if customer order *j* is delivered from distribution center *m* by using transportation mode *k*, 0 otherwise
PEmjk	Probability of earliness if customer order *j* is delivered from distribution center *m* by using mode *k*
Imo	Starting inventory level in distribution center *m*
Im	Ending inventory level in distribution center *m*.

3.2 Model

Objective Function:

$$\text{Max}\, z = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \tag{1}$$

Subject to:

$$\sum_{m=1}^{NDC}\sum_{k=1}^{NK} X_{mjk} = 1 \qquad j = 1, 2, \ldots, NC \tag{2}$$

$$Im = Imo - \sum_{j=1}^{NC}\sum_{k=1}^{NK} X_{mjk} Q_j \qquad m = 1, \ldots, NDC \tag{3}$$

$$PE_{mjk} = p\left(z \le \left(\frac{D_j - \mu_{mjk} X_{mjk}}{\sigma_{mjk} X_{mjk}}\right)\right)$$
$$m = 1, \ldots, NDC,\ j = 1, \ldots, NC, k = 1, \ldots, NK \tag{4}$$

$$PE_{mjk} \le X_{mjk} \qquad m = 1, \ldots, NDC,\ j = 1, \ldots NC, k = 1, \ldots, NK \tag{5}$$

$$Z1 = \sum_{m=1}^{NDC}\sum_{j=1}^{NC}\sum_{k=1}^{NK} t_{mjk} X_{mjk} \qquad m = 1, \ldots, NDC,\ j = 1, \ldots NC, k = 1, \ldots, NK \tag{6}$$

$$Z2 = \sum_{m=1}^{NDC}\sum_{j=1}^{NC}\sum_{k=1}^{NK} p_{mjk} \qquad m = 1, \ldots, NDC,\ j = 1, \ldots NC, k = 1, \ldots, NK \tag{7}$$

$$Z3 = \sum_{m=1}^{NDC}\sum_{j=1}^{NC}\sum_{k=1}^{NK} CE_{mjk} X_{mjk} \qquad m = 1, \ldots, NDC,$$
$$j = 1, \ldots NC, k = 1, \ldots, NK \tag{8}$$

$$Z4 = \min(I_m) \qquad m = 1, \ldots, NDC \tag{9}$$

$$\lambda_1 = 1 - \frac{(z1 - lb1)}{(ub1 - lb1)} \tag{10}$$

$$\lambda_2 = \frac{(z2 - lb2)}{(ub2 - lb2)} \tag{11}$$

$$\lambda_3 = 1 - \frac{(z1 - lb1)}{(ub1 - lb1)} \tag{12}$$

$$\lambda_4 = \frac{(z4 - lb4)}{(ub4 - lb4)} \tag{13}$$

$$X_{mjk} \in \{1, 0\} \qquad m = 1, \ldots, NDC,\ j = 1, \ldots, NC, k = 1, \ldots, NK \tag{14}$$

$$0 \leq P_{mjk} \leq 1 \qquad m = 1, \ldots, NDC, j = 1, \ldots, NC, k = 1, \ldots, NK \tag{15}$$

The objective function (1) seeks to maximize the satisfaction level for the four objectives. Constraint (2) ensures that each customer demand is met. Constraint (3) calculates the ending inventory in each distribution center, which is equal to starting inventory minus shipment made from the distribution center to customers. Constraints (4) and (5) captures the probability of earliness. The values of performance measures are determined by constraints (6)–(9) for transportation cost, total probability, carbon emission and inventory level respectively. Constraints (10)–(13) calculates the satisfaction levels to the various objective values and the desired upper and lower bound. Constraint (14) is a binary constraint and (15) ensures the probability of earliness is between zero and one.

4 Numerical Experiment

Experiments are conducted in this section to validate the model but limited due to space restrictions. We implement the model in LINGO/WIN64 19.0.32 and solve various instances of the model. Data is generated with 10 customers and 2 Distribution Centers in our first experiment, the various customers a DC will serve is determined as shown in Fig. 2. The various probability of on-time delivery is also included, and results displayed.

In the other experiments, varying leadtimes are considered to determine the optimal route as well as the particular DC to order from. As shown in Table 1, '1' indicates if a particular route is used and the corresponding DC for the various instances, and '0' otherwise.

Table 2 gives a summary of the fuzzy parameters for the various instances with Table 3 presenting the beginning and ending inventory levels for the start and end of the experiment. We have used different weights for satisfaction levels using one-dominant and three-equal scheme with two different values (0.4 and 0.7) as shown in

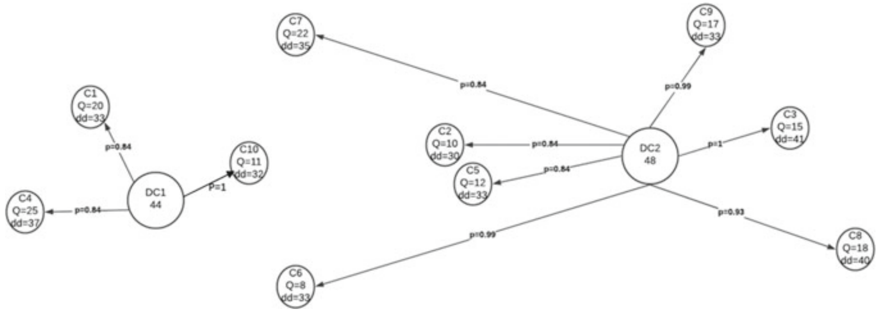


Fig. 2 Distribution with probability of on-time delivery

Table 1 Varying leadtimes calculations

Customers	Shorter leadtimes				Mixed leadtimes				Long leadtimes			
	DC 1		DC 2		DC 1		DC 2		DC 1		DC 2	
	Train	Truck	Train	Truck	Train	Truck	Train	Truck	Train	Truck	Train	Truck
1	0	0	0	1	0	1	0	0	1	0	0	0
2	0	0	1	0	0	0	0	1	0	0	1	0
3	0	1	0	0	1	0	0	0	0	1	0	0
4	0	0	0	1	0	0	1	0	0	0	1	0
5	0	0	1	0	0	0	0	1	1	0	0	0
6	0	0	1	0	0	1	0	0	1	0	0	0
7	1	0	0	0	0	0	0	1	0	0	1	0
8	1	0	0	0	0	0	1	0	0	0	1	0
9	1	0	0	0	0	0	1	0	0	0	1	0
10	1	0	0	0	0	1	0	0	0	0	1	0

Table 2 Fuzzy parameter values and corresponding objective values

Fuzzy parameters	Short	Mixed	Long
λ1	0.97 0.97	0.90 0.90	0.86 0.84
λ2	0.99 0.99	0.99 0.99	0.89 0.92
λ3	0.93	0.74	0.55
λ4	1.00	1.00	0.99 0.99
Z	3.91 2.97	3.63 2.89	3.31 2.76

Table 4. The bold values show the higher weight for the corresponding performance measures for the mixed leadtime case.

Table 3 Begining and ending inventory

Beginning inventory			
I1	200		
I2	250		
Ending inventory			
	Short	Mixed	Long
I1	146 146	146 146	145
I2	146 146	146 146	147

Table 4 Satisfaction level with weights

	{0.4, 0.2, 0.2, 0.2}				{0.7, 0.1, 0.1, 0.1}			
	C1	C2	C3	C4	C5	C6	C7	C8
λ1	**0.90**	0.90	0.90	0.90	**0.91**	0.89	0.90	0.90
λ2	0.99	**0.99**	0.99	0.99	0.97	**0.99**	0.99	0.99
λ3	0.74	0.74	**0.76**	0.73	0.75	0.72	**0.76**	0.74
λ4	1	1	0.95	**1**	0.97	0.98	0.95	**1**

5 Results and Conclusion

The results of the experiment carried out are summarized in Tables 1, 2, 3 and 4. Based on these results we can conclude that train was the preferred mode when the distances were long (9/10 times). However, the use of distribution centers was relatively balanced in all cases. All satisfaction levels were high in all three cases except carbon emission, and as expected, the satisfaction level dropped significantly as the distances increased. Other satisfaction levels slightly dropped when long-distance case was considered. Satisfaction level measuring carbon emission was more sensitive to the weights than other measures as shown in Table 4.

We are planning to extend this model to consider multiple items and multiple periods with more customers and distribution centers in the future works.

References

1. Azizi, V., Hu, G.: Multi-product pickup and delivery supply chain design with location-routing and direct shipment. Int. J. Prod. Econ. **226**, 107648 (2020)
2. Benjamin, J.: An analysis of mode choice for shippers in a constrained network with applications to just-in-time inventory. Transp. Res. Part B **24**(3), 229–245 (1990)
3. Chan, E.T.S., Chung, S.H.: Multi criteria genetic optimization for distribution network problems. Int. J. Adv. Manuf. Technol. **24**, 517–532 (2004)
4. Chiang, W.C., Russell, R.A.: Integrating purchasing and routing in a propane gas supply chain. Eur. J. Oper. Res. **154**, 710–729 (2004)
5. Chopra, S., et al.: Supply Chain Management: Strategy, Planning, & Operation. Stanford University (2007)
6. Contesses, L., Ferrer, J., et al.: A mixed integer programming model for gas purchase and transportation. Ann. Oper. Res. **139**, 39–63 (2005)
7. Cosma, O., Dănciulescu, D., Pop, P.C.: On the two-stage transportation problem with fixed charge for opening the distribution centers. IEEE Access **7**, 113684–113698 (2019)
8. Diabat, A., Dehghani, E., Jabbarzadeh, A.: Incorporating location and inventory decisions into a supply chain design problem with uncertain demands and lead times. J. Manuf. Syst. **43**, 139–149 (2017)
9. Farahani, R.Z., Elahipanah, M.: A genetic algorithm to optimization the total cost and service level for just-in-time distribution in a supply chain. Int. J. Prod. Econ. **111**(2), 229–243 (2008)
10. Hsiao, Y.H., Chen, M.C., Chin, C.L.: Distribution planning for perishable foods in cold chains with quality concerns: formulation and solution procedure. Trends Food Sci. Technol. **61**, 80–93 (2017)

11. Humair, S., Ruark, J.D., Tomlin, B., Willems, S.P.: Incorporating stochastic lead times into the guaranteed service model of safety stock optimization. Interfaces **43**(5), 421–434 (2013)
12. Kumar, R., Ganapathy, L., Gokhale, R., Tiwari, M.K.: Quantitative approaches for the integration of production and distribution planning in the supply chain: a systematic literature review. Int. J. Prod. Res. **58**(11), 3527–3553 (2020)
13. Lee, Y.H., Kim, H.S.: Production-distribution in supply chain considering capacity constraints. Comput. Ind. Eng. **43**, 169–190 (2002)
14. Mohammed, A., Wang, Q.: The fuzzy multi-objective distribution planner for a green meat supply chain. Int. J. Prod. Econ. **184**, 47–58 (2017)
15. Molla-Alizadeh-Zavardehi, S., Shoja, A.: A fuzzy multi-product two-stage supply chain network design with possibility of direct shipment. Ann. Optim. Theory Pract. **1**(2), 11–22 (2018)
16. Puga, M.S., Minner, S., Tancrez, J.S.: Two-stage supply chain design with safety stock placement decisions. Int. J. Prod. Econ. **209**, 183–193 (2019)
17. Schulz, E.P., Diaz, M.S., et al.: Supply chain optimization of large scale continuous processes. Comput. Chem. Eng. **29**, 1305–1316 (2005)
18. Tang, X., Lehuédé, F., Péton, O.: Location of distribution centers in a multi-period collaborative distribution network. Electron. Notes Discrete Math. **52**, 293–300 (2016)
19. Zhou, G., Min, H., et al.: The balanced allocation of customers to multiple distribution centers in the supply chain network: a genetic algorithm approach. Comput. Ind. Eng. **43**, 251–261 (2002)

Business Analysis of Electronic Device Manufacturers on Business Continuity Plans Under Uncertain Supply Chain Disruption Risks

Kotomichi Matsuno, Noriyuki Hosokawa, and Takahiro Ohno

Abstract Since the Great East Japan Earthquake in 2011, the importance of BCP has been recognized widely, as a result, formulation rate of BCP in Japanese manufacturing industry has increased to about 50%. However, nearly 80% of them are major manufacturers. There is a difficult task on supporting small and medium-sized manufacturers in developing BCP. In addition, in recent years, supply chains have become much more complex increasingly due to the globalization. Not only large-scale supply chain disruptions such as natural disasters or industrial accidents, but also local and small-scale disruptions such as power outages or information system failures can cause enormous losses. Even if a BCP is formulated, whether it is an appropriate BCP or not has not been sufficiently discussed quantitatively. To support the manufacturers to formulate an appropriate BCP, in this study, a stochastic model of supply chain disruptions and a BCP model are created based on a supply chain consists of component suppliers, retailers and an electronics device manufacturer as the main participant. Loss mitigation effects by different levels of BCP are analyzed to help the manufacturer to formulate appropriate BCP according to the attributes by numerical simulation.

Keywords Business continuity plan (BCP) · Supply chain disruption risks · Resilience

1 Introduction

Owing to disasters such as Hurricane Katrina 2005, Great East Japan Earthquake 2011, Thailand Flood 2011, Kumamoto Earthquake 2016, and the recent Covid-19, supply chains have suffered from heavy losses. Ho et al. [1] presented that supply chain disruptions cause significant losses to countless manufacturers, and make manufacturers get difficult in surviving in the market. Since the Great East Japan Earthquake in 2011, the importance of BCP has been recognized widely in

K. Matsuno (✉) · N. Hosokawa · T. Ohno
Waseda University, Tokyo, Japan
e-mail: sidiwu@aoni.waseda.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_35

Japan. As a result, in the recent decade, the rate of BCP formulation and implementation in Japanese manufacturing industry has increased to about 50%. On the other hand, creating and keeping a more competitive supply chain is considered as the basis of business. So, risk hedging and recovery strategies decision-making with proper effectiveness analysis and evaluation is considered as the most important process in BCP formulation.

Products such as LCD TVs and PCs in Japan market have already entered the stable period after the introduction and competition period. In competition period, according to commoditization of key components and production capacity oversupply, intense price competition occurs among manufacturers. As a result, it is difficult for most of the manufacturers to achieve enough profit to get survived in the market. For these manufacturers with extremely low profit rate, a highly efficient BCP based on proper risk mitigation measures is necessary. Figure 1 shows the supply chain which consists of multiple component suppliers such as LCD panel manufacturers, motherboard manufacturers, power supply unit manufacturers and product manufacturer and multiple retailers. Product manufacturer assembles products with the components and sells products to retailers.

In this study, profit of the electronic device manufacturer is analyzed to clarify the impact on components procurement disruptions caused by disasters of medium-large scale earthquakes to a supply chain with long component procurement lead time. As a risk hedging and recovery countermeasure, multiple sourcing as robust countermeasure and multiple sourcing with production capacity reservation as robust and flexible countermeasure are proposed to verify the loss aversion effect. Different situation such as related areas of procurement activity, inventory level of components is also discussed according to the numerical experiments.

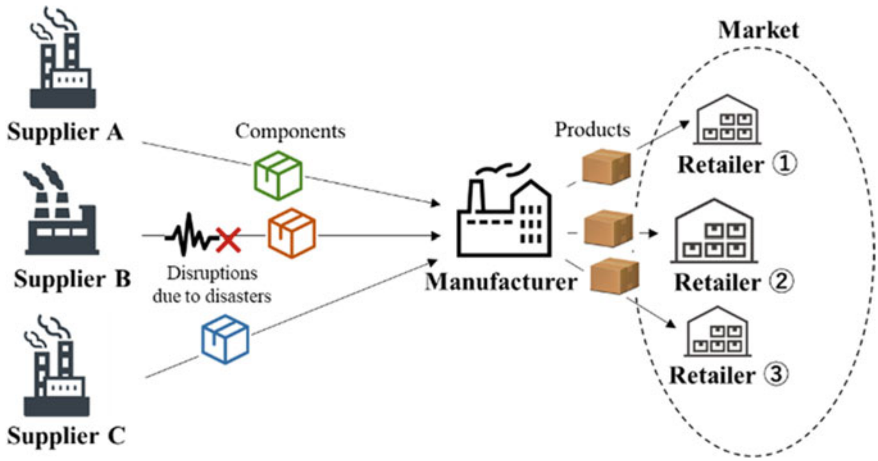


Fig. 1 Supply chain structure and component procurement disruptions due to disasters

2 Literature Review

Many articles discussed the impact of disasters on supply chain disruptions as well as risk hedging and recovery countermeasures for supply chain participants since Great East Japan Earthquake 2011. Tokui et al. [2] examined the economic impact of supply chain disruptions immediately after the Great East Japan Earthquake and estimated the amount of production losses caused by the supply chain disruptions to at least 0.35% of the country's GDP. Park et al. [3] discussed the Japanese manufacturing firms about their responses to the recent natural disasters such as earthquake, tsunami, and nuclear disasters. Based on case studies, several supply chain recovery processes after natural disasters as well as humanitarian disruptions were discussed and reflection points in terms of disaster planning and recovery responses were summarized.

Except for articles based on qualitative analysis, surveys, etc., many articles focusing on the analysis of the effects of preventive and recovery countermeasures against supply chain disruptions caused due to disasters. Amanda [4] analyzed and confirmed the effect on reducing supply chain risks with different inventory placement and back-up methodologies in a multi-echelon network. Both supply disruption risks and demand uncertainty were taken into consideration to run simulation to verify the effectiveness of risk mitigation. Masuda et al. [5] built several supply chain structural models and discussed the tradeoff relation between supply chain efficiency and robustness under supply chain disruption risks. Matsukawa et al. [6] visualized supply chains and studied countermeasures to mitigate supply chain disruptions risks in terms of redundancy, robustness, and flexibility. Chen et al. [7] focused on supply chain disruptions caused by Covid-19. Since it is a global pandemic, manufacturers are difficult to get recovered from their supply chain disruptions in a short period. A supply chain disruption preventive strategy was proposed by changing original product types. Also, a mixed integer linear programming model was developed with combining emergency procurement on the supply side and product type changing by the manufacturer as well as the demand side to maximize the total profit.

In this study, as a robust and flexible countermeasure for risk hedging and recovery, capacity reservation (CR) for manufacturer is proposed based on multiple sourcing procurement. Effectiveness reducing opportunity loss due to procurement disruptions is to be confirmed. More related areas of a supply chain, higher disruption probability will be. Number of related areas of procurement activity which is one of the attributes of supply chain is considered as a decision variable to the probability of procurement disruptions. Moreover, inventory level of components which can be considered as another attribute of supply chain is taken to analyze the effectiveness of the proposal.

3 Procurement Disruptions and Supply Chain Model

3.1 Procurement Disruptions and Recovery Periods

Procurement disruptions due to earthquake in Japan are focused to create the disruption model in this study. Occurrence probability of earthquakes and recovery period from earthquakes are key elements which should be decided during model creating. With the data about earthquakes over seismic intensity 6.5 in the last 10 years by Japan Weather Association in Table 1, average annual probability of an earthquake which seismic intensity is 7 could be estimated. Since seismic intensity 7 earthquakes occurred 4 times in 11 years including 2021, average annual occurrence probability is 0.36 and average monthly occurrence probability is calculated as 0.03.

Table 2 shows the recovery periods of several electronic device manufacturers after the Kumamoto earthquake in 16th Apr. 2016. Date of partial recovery in the table is the date when production was restarted partially. Date of full recovery in the table is the date when production was recovered as the same as the capacity before earthquake. From Table 2, we can find that most of the average recovery periods was nearly 1 month after the earthquake.

Based on the above information about the occurrence probability of seismic intensity 7 earthquake and average recovery period of electronic device manufacturers. Average monthly occurrence probability of object earthquakes is set to 0.03/month in Japan and average recovery period of electronic device manufacturers is set to 1 month for running simulation.

Table 1 Earthquakes over seismic intensity 6.5 in the last 10 years in Japan [8]

Date and time	Epicenter	Magnitude	Max seismic intensity
2021/2/13 23:08	Coast of Fukushima	M7.3	Upper 6
2019/6/18 22:22	Coast of Yamagata	M6.7	Upper 6
2018/9/6 3:08	Eastern Iburi of Hokkaido	M6.7	7
2016/4/16 3:55	Aso region of Kumamoto	M5.8	Upper 6
2016/4/16 1:25	Kumamoto region of Kumamoto	M7.3	7
2016/4/15 0:03	Kumamoto region of Kumamoto	M6.4	Upper 6
2016/4/14 21:26	Kumamoto region of Kumamoto	M6.5	7
2011/4/7 23:32	Coast of Miyagi	M7.4	Upper 6
2011/3/15 22:31	Western Shizuoka	M6.0	Upper 6
2011/3/12 3:59	Chuetsu region in Niigata	M6.6	Upper 6
2011/3/11 14:46	Sanriku offshore	M7.9	7

Table 2 Recovery periods of electronic device manufacturers after Kumamoto earthquakes 2016 [9]

Business office	Location	Date of partial recovery	Date of full recovery
Renesas Semicond. Manuf. Co., Ltd., Kawashiri factory	Kumamoto-shi	Apr. 22, 2016	May 22, 2016
FUJIFILM Kyusyu Co., Ltd	Kikuyomachi	Apr. 23, 2016	May 23, 2016
Mitsubishi Electric Corp. Power Device Works Kumamoto	Koushi-shi	May 9, 2016	May 31, 2016
Tokyo Electron Kyushu Ltd	Koushi-shi	Apr. 25, 2016	End of June, 2016
Sony Semicond. Manuf. Co., Ltd., Kumamoto Tech. Ctr	Kikuyomachi	May 9, 2016	End of July, 2016

3.2 Supply Chain Model

As introduced in Sect. 1, product manufacturer assembles products with the components procured from suppliers and sells products to the retailers. In order the make the analysis simpler, a supply chain consists of a supplier, a manufacturer and a retailer which is shown in Fig. 2 is to be discussed. Comparing with order receiving cycle of products, procurement lead time of electronic devices components such as LCD panel is much longer. With reference to some cases, order receiving cycle of products is set to 1 month, lead time of component procurement is set to 3 months.

Profit of manufacturer is considered as an indicator to clarify the impact on components procurement disruptions as well as effectiveness of risk hedging and recovery countermeasures. Formula (1) shows the profit of manufacturer which is the difference between sales amount and costs including component procurement cost, inventory cost and production cost.

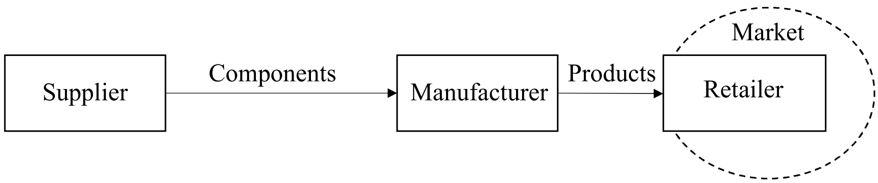


Fig. 2 A simplified supply chain structure based on Fig. 1

Table 3 Disruption probability for component procurement under different numbers of related areas

Numbers of related areas	1	2	3	4	5	6	7
Disruption probability	0.0043	0.0087	0.0130	0.0173	0.0216	0.0260	0.0303

$$\pi_m = \sum_{n=1}^{N}\Big\{p_p \min\big(I_{n-1}, q_{p.n}\big) - \alpha c_{proc.} q_{c.n} lt_{proc.} - \frac{1}{2} c_{inv.} q_{c.n} lt_{proc.} - \big[c_f + c_v \min\big(I_{n-1}, q_{p.n}\big)\big]\Big\} \quad (1)$$

p_p Product price
$q_{p.n}$ Volume of product order from retailer
I_n Inventory level of components
$c_{proc.}$ Component procurement cost (per unit)
$q_{c.n}$ Volume of component procurement
$lt_{proc.}$ Component procurement lead time
$c_{inv.}$ Inventory cost (per unit)
c_f Fixed cost of production
c_v Variable cost of production
α Coefficient for determining procurement timing. If there is no remssainder of $(n-1)/lt_{proc.}$ $\alpha = 1$ otherwise $\alpha = 0$.

Moreover, related area is an important attribute of supply chain. Japan consists of 7 regions which are used for statistical by government officials. They are Hokkaido, Tohoku, Kanto, Chubu, Kinki, Chugoku-Shikoku and Kyushu-Okinawa. In this study, relation between number of related areas of component procurement activities and probability of disruptions is focused to verify the disruption impact. Table 3 shows the relation which is calculated with number of regions in Japan and average monthly occurrence probability of object earthquakes discussed in 3.1.

4 Models of Risk Hedging and Recovery Countermeasures

4.1 Robust Countermeasure

Generally, multiple sourcing is considered as a robust countermeasure for hedging disruption risks. Figure 3 shows a simple robust structure with one more supplier adding to Fig. 2. Profit of manufacturer in this structure is shown in Formula 2.

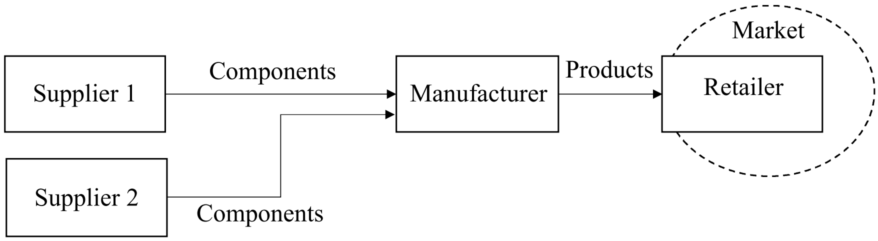


Fig. 3 A simplified robust supply chain structure based on Fig. 2

$$\pi_m = \sum_{n=1}^{N} p_p \min\left(I_{n-1}, q_{p.n}\right) - \alpha c_{proc.} lt_{proc.}(q_{c1.n} + q_{c2.n}) - \frac{1}{2} c_{inv.} lt_{proc.}(q_{c1.n} + q_{c2.n}) - \left[c_f + c_v \min\left(I_{n-1}, q_{p.n}\right)\right] \tag{2}$$

$q_{c1.n}, q_{c2.n}$: Volume of component procurement from supplier 1 and supplier 2.

4.2 Robust and Flexible Countermeasure

Supply chain coordination with capacity reservation (CR) was discussed in many articles such as Cheaitou et al. [10], Serel et al. [11]. In this study, CR is considered as a flexible countermeasure for reducing opportunity losses and keeping component inventory in a low level. Comparing with the robust supply chain discussed in 4.1, in the event of a disruption in component procurement from one supplier, it is possible to achieve more volume of components with CR.

Figure 4 shows the cost structure with using CR. Manufacturer should make a lump-sum payment according to upper limit of capacity to supplier to reserve production capacity in advance. Components can be purchased by normal procurement fee if the volume of components is less than upper limit of purchase volume which was decided beforehand. More components over upper limit of purchase volume can be purchased by slightly higher price if the manufacturer decide to utilize reserved production capacity. Figure 5 shows a simple robust and flexible supply chain structure with CR adding to Fig. 3. Profit of manufacturer in this structure is shown in formula 3.

$$\pi_m = \sum_{n=1}^{N} p_p \min\left(I_{n-1}, q_{p.n}\right) - 2l s_{CR} - \alpha\left[c_{proc.} lt_{proc.}(q_{c1.n} + q_{c2.n}) + c_{cr} lt_{proc.}(q_{cr1.n} + q_{cr2.n})\right] - \frac{1}{2} c_{inv.} lt_{proc.}(q_{c1.n} + q_{cr1.n} + q_{c2.n} + q_{cr2.n}) - \left[c_f + c_v \min\left(I_{n-1}, q_{p.n}\right)\right] \tag{3}$$

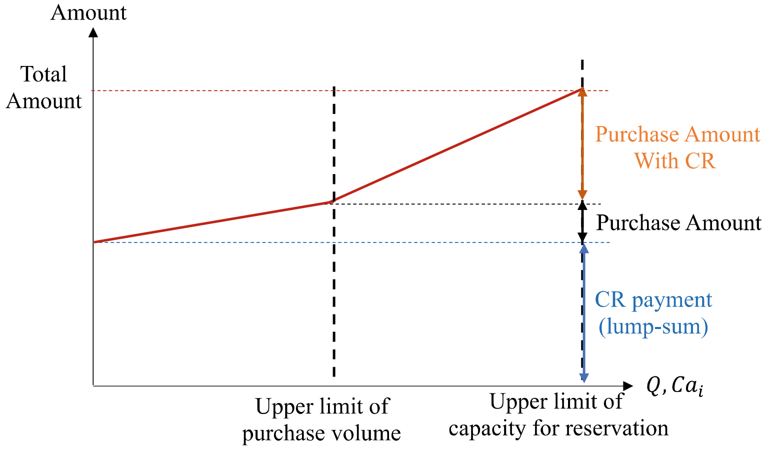


Fig. 4 Cost structure of capacity reservation (CR)

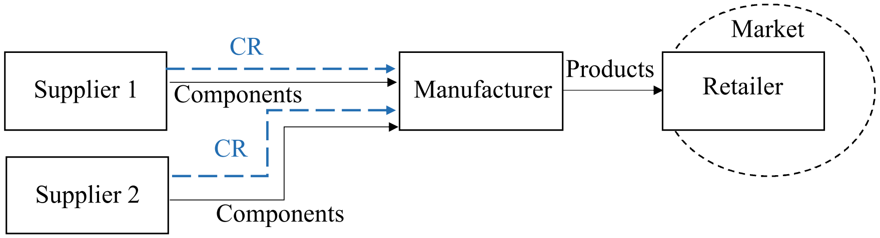


Fig. 5 A simplified robust and flexible supply chain structure based on Fig. 3

$q_{cr1.n}, q_{cr2.n}$	Volume of component procurement with using reserved capacity from supplier 1 and supplier 2
c_{cr}	Component procurement cost (per unit) to $q_{cr1.n}$ and $q_{cr2.n}$
ls_{CR} :	Lump-sum payment to one supplier for capacity reservation

5 Numerical Experiments and Results

Supply chain attributes including related areas of procurement and component inventory level of manufacturer are considered to analyze the profit of manufacturer to clarify the loss aversion effect among simple model, robust model, robust and flexible model. Period of experiments is set to 3 years, and data for experiments is shown as below (Table 4).

Figure 6 shows total profit of manufacturer which is the average value of 50 simulation runs with different number of related areas. Vertical axis is the ratio of profit

Table 4 Numerical experiments data

Est. product demand (month): 1000	Component proc. cost: 10,000	Component order lot size: 3000
Production cost (fixed): 10,000,000	Production cost (variable): 1000	Component proc. LT: 3 months
Act. product demand (month): $N\left(1000, 200^2\right)$	Product price: 35,000	Inventory cost (month): 200

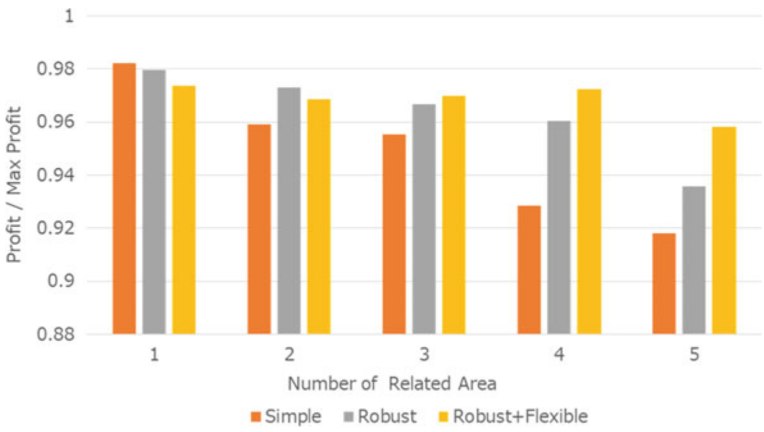


Fig. 6 Relation between numbers of related areas and total profit of manufacturer

to maximum profit which could be obtained in simple model without any disruptions in 3 years. More related areas of procurement, lower profit of manufacturer in simple model would be. With robust and flexible model, no significant reduction of profit has been confirmed with increasing in related areas. Figure 7 shows total profit of manufacturer under 5 procurement related areas relation with different component inventory level. In a high disruption probability procurement environment, effectiveness of a slight increase in inventory level in simple model is confirmed. It is better to keep a lower inventory level due to the reserved capacity by robust and flexible model.

6 Conclusions

Three supply chain models by two supply chain attributes including number of procurement related areas and component inventory level are discussed in this study. The proposed robust and flexible model shows the best performance in losses aversion under procurement environment of large number related areas. Moreover, operation efficiency in normal times can also be kept by low inventory level with the proposed model. For the manufacturers that discussed in Sect. 1, the conclusions of this study can be used as reference for decision-making in BCP formulation.

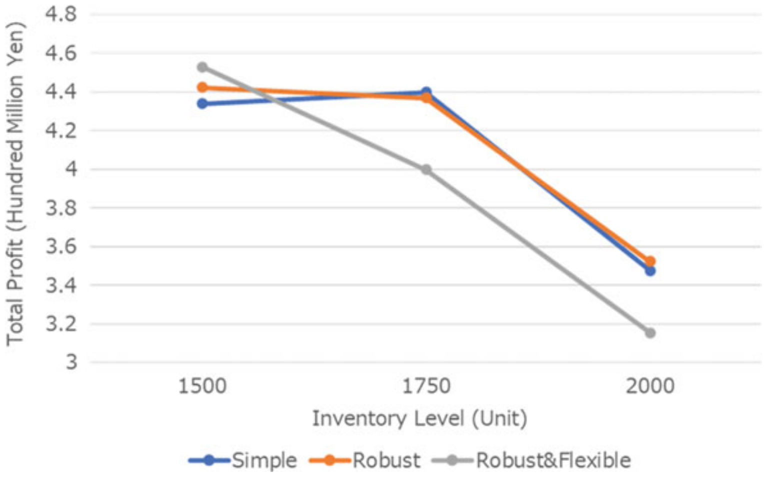


Fig. 7 Relation between component inventory level and total profit of manufacturer under 5 procurement related areas

References

1. Ho, W., Zheng, T., Yildiz, H., Talluri, S.: Supply chain risk management: a literature review. Int. J. Prod. Res. **53**(16), 5031–5069 (2015)
2. Tokui, J., Kawasaki, K., Miyagawa, T.: The economic impact of supply chain disruptions from the Great East-Japan earthquake. Jpn. World Econ. **41**, 59–70 (2017)
3. Park, Y., Hong, P., Roh, J. J., J. R.: Supply chain lessons from the catastrophic natural disaster in Japan. Bus. Horiz. **56**, 75–85 (2013)
4. Schmitt, A.J., Singh, M.: A quantitative analysis of disruption risk in a multi-echelon supply chain. Int. J. Prod. Econ. **139**(1), 22–32 (2012)
5. Masuda, T., Mizuno, H.: Study for risk management of disruption in supply chains. Proc. Sch. Inf. Telecommun. Eng. Tokai Univ. **10**(1), 25–35 (2017)
6. Matsukawa, H., Liu, T.: A supply chain visualization system and its validation analysis. J. Jpn. Ind. Manag. Assoc. **67**(2), 73–82 (2016)
7. Chen, J., Wang, H., Zhong, R.Y.: A supply chain disruption recovery strategy considering product change under COVID-19. J. Manuf. Syst. Press (2021). https://doi.org/10.1016/j.jmsy.2021.04.004
8. Japan Weather Association: Homepage. https://earthquake.tenki.jp/bousai/earthquake/entries/level-6-plus/
9. Kashima, H.: The 2016 Kumamoto earthquake and manufacturing industry: regional characteristics on the disasters and restoration processes. Ann. Assoc. Econ. Geogr. **64**(2), 138–149 (2018)
10. Cheaitou, A., Cheaytou, R.: A two-stage capacity reservation supply contract with risky supplier and forecast updating. Int. J. Prod. Econ. **209**(C), 42–60 (2019)
11. Serel, D.A., Dada, M., Moskowitz, H.: Sourcing decisions with capacity reservation contracts. Eur. J. Oper. Res. **131**(3), 635–648 (2001)

A Vision for a Highly Automated Digital Local Manufacturing Network—Solutions and Challenges

Mikael Collan, Jyrki Savolainen, Veli-Matti Virolainen, and Pasi Luukka

Abstract There is a lot of evidence of the manufacturing strategy shift from mass production to the individual and decentralized micro-production, which in essence means a shift from low-cost country production to production at the place of consumption. This paper presents a vision for a highly automated digital urban manufacturing network, based on a central market mechanism for matching designers, customers, and producers operating digital-manufacturing machinery. The mechanism proposed automatically clears the market by matching supply and demand, based on pre-given or auction-based bids with an optimal allocation. Contracting, profit-sharing and settlement are automated according to pre-specified market rules. Network participant profiles are used and updated to establish reputation and trust that are used in the allocation optimization. This paper presents the elements and logic of an automated digital urban manufacturing network on a technical level and discusses some of the benefits reached by shifting centralized production to decentralized urban networks in a wider perspective. Practical implications to the real-world industries, where the presented vision is relevant are shortly discussed.

Keywords Digital manufacturing · Platformization · Automatic contracting · Optimal allocation · Digital twin

M. Collan · J. Savolainen · V.-M. Virolainen (✉) · P. Luukka
LUT-University, P.O. Box 20, 53851 Lappeenranta, Finland
e-mail: veli.matti.virolainen@lut.fi

M. Collan
e-mail: mikael.collan@vatt.fi

J. Savolainen
e-mail: jyrki.savolainen@lut.fi

P. Luukka
e-mail: pasi.luukka@lut.fi

M. Collan
VATT Institute for Economic Research, P.O. Box 1279, 00101 Helsinki, Finland

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_36

1 Introduction

Industry 4.0 can be described shortly as a new digitally-driven manufacturing paradigm that is based on the efficient use of smart machines steered by massive amounts of data [1]. According to Li [2], in the center of Industry 4.0 is the integration of new technologies such as Additive Manufacturing (AM), cloud computing, Internet of Things (IoT) with people and with products. Four important disruptions that face the manufacturing industry can be summarized as: (i) the increase of data volumes; (ii) analytics and business intelligence; (iii) human–machine interaction; and (iv) the ability to transfer digital information to the physical world [3]. In vein with the industry 4.0 paradigm (or manufacturing 4.0) that follows from it, this paper sets out to discuss manufacturing systems that are highly digitalized and that are based on a high level of digital automation. The discussed systems reside within the areas of the second and the fourth types of disruption identified.

Within the realm of highly digitalized manufacturing systems and network of systems, each machine- or system-instance can be understood as a unique cyber-physical system (CPS) that has both a physical part and a virtual-counterpart that exists as a node in "cyberspace"—in other words, the tangible machines are mapped in a virtual model-space that maps the physical reality with high fidelity and enables the fully digital and automatic (machine-initiated) information flow between system nodes. This refers to the type of constructions, often enhanced with real-time data connections and asset-specific life-cycle data-management, referred to as Digital Twins [4]. Typically, digital twins depict equipment inside single factory premises and are in place to form efficient cyber-physical production systems (CPPS) by combining separate CPS-entities into a system model with the goal of operating the system with high technical and economic efficiency. Most often the complexity in the modern digital twins is reduced by the static nature of the set-ups—the system configuration changes are kept to a minimum or are not allowed. According to Oztemel and Gursev [5], there are already, multi-company initiatives, such as Industrial Internet Consortium from 2014, which aim to bring together operational systems and digital technology to the widest possible extent.

Now, we leverage the thinking behind the digital twin and combining cyber-physical systems in the context of manufacturing 4.0 to the broader view of the inter-operation between digitally operating agents that represent the demand and supply of manufacturing goods and the presented envisioned case is that of a fully digital market mechanism for products manufactured directly from digital blueprints. The vision differs from what can be already found existing through "hubs", ideas of fully automated "dark" factories (see, e.g., [5]) or digital marketplaces/platforms for additive manufacturing in that the operations are conducted without human involvement in the operational phase (while the system allows for manual access also) between digitally operating machines through the digital market mechanism. The point of reference here is the concept of "local manufacturing" or "urban manufacturing", which refers to fleets of digital manufacturing facilities operating near customers and not necessarily within "factory" environments. See Fig. 1 for a high-level sketch of

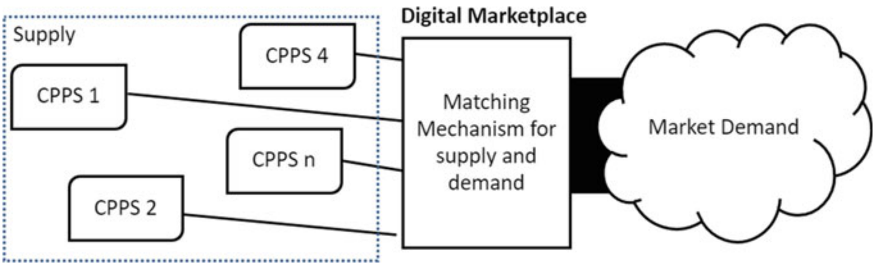


Fig. 1 A high-level illustration of the digital market mechanism

a digital market mechanism. As the demand is digitally managed and, if demanded, routed to the production nodes closest to the demand there are consequences not only to manufacturing and to matching supply with demand, but also, and importantly, to supply chains and to what can be called "market positions" of agents in both the supply and the demand side. Importantly digital platformization of manufacturing may drive market power to the platforms in a similar way that has happened with the well-known platform businesses that have "captured" the retail markets of many standard goods internationally. Such platforms will disrupt also the value chains and the accumulation of value between the producers and the platform and thus will affect also the global manufacturing value chains (GVCs).

The change brought by the change to local production and smaller manufacturing units has been acknowledged in previous literature. Despite the ongoing technical progress in the individual machine-to-machine communication protocols such as OPC-UA [6], the literature is lacking accounts on how to coordinate such systems in a way that would, even in theory, come close to the economics provided by the current paradigms of mass manufacturing. Here we address the question of how costs of networked small-scale manufacturing units can be brought down by proposing a centralized and automated digital market mechanism that can be used in efficient and dynamic coordination of a network consisting of multiple production nodes operating fully digitally and treating the network as a source of a large production capacity of goods. If the digital coordination is highly efficient the difference between single node high capacity and multiple node low-capacity network is made smaller.

Täuscher and Laudien [7] define marketplaces as platforms that enable and support transactions between independent supply and demand-side participants. They observe that the winner-takes-all dynamics of marketplaces have attracted hundreds of firms to launch their own marketplaces. The research on eCommerce and its´ many forms are already active for more than three decades and a community of researchers in the Information Systems Science domain has contributed to the topic. While the "talk" is about a market-mechanism, or a market-place, the same issues have also been discussed in the supply chain literature. According to Mentzer et al. [8] supply chain can be defined as a set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and/or information from a source to a customer. They have further identified three degrees of supply chain complexity: a “direct supply chain,” an “extended supply chain,” and an “ultimate supply chain.” An ultimate supply chain includes all the organizations involved in all the upstream and downstream flows of products, services, finances, and information from the ultimate supplier to the ultimate customer. When we discuss a fully digital matching of supply and demand, for the sake of simplicity, we leave out the supply chain issue, while it underlies the production side of the equation. On a further note, it has been understood for a long time that global supply chains are in fact dynamic networks of interconnected firms and industries—this notion rings a familiar tone with what is proposed here.

Analytics in this context can be categorized as Big Data Business Analytics (BDBA) which will provide means to gain value from ever-growing massive amount of data [9]. With big data, analysis of the data is concentrating on the following qualities: velocity, variety, and volume [10]. This is then combined with business analytics where analysis is more concentrating on descriptive, predictive, and prescriptive analytics [11]. In manufacturing context with BDBA one can improve the visibility, flexibility, and integration of CPPS networks. Predictive and prescriptive analytics are in a key role in guiding companies to make solid decisions on the strategic directions [12]. Through BDBA we can evaluate CPPS networks through evidence-based data, predictive forecasting, mathematical modelling, operational and statistical analysis and through simulation and optimization techniques [9]. The systems required for the envisioned mechanism include the smart ability of splitting and matching demand with supply in an optimized way—the problem may seem trivial, but the difficulty grows with the number of network nodes included and if large orders must be split between multiple production nodes.

All in all the next presented vision is a product of combining the real-world digitally operated manufacturing units with the concept of digital twins brought to the context of clearing supply and demand within a concentrated market mechanism that can operate quasi-autonomously. As so far always, human is in the loop in that most decision-making, pre-approved or not, has at some point landed at a human desk. Next, we present the vision, then some conclusions are drawn and future research directions discussed.

2 Vision for an Automated Digital Local Manufacturing Network

The vision presented here is based on a marketplace-mechanism that intelligently and automatically matches the demand for local manufacturing with the supply. The type of products that are described are standard and ready-made designs with material specifications included, for example spare parts for existing machinery within a given context (industry). The vision is based on the notion of having a digital product store that lists, with pricing and other information the products available for purchase (Design store/depository in Fig. 2). The demand is of the type “product + quantity”

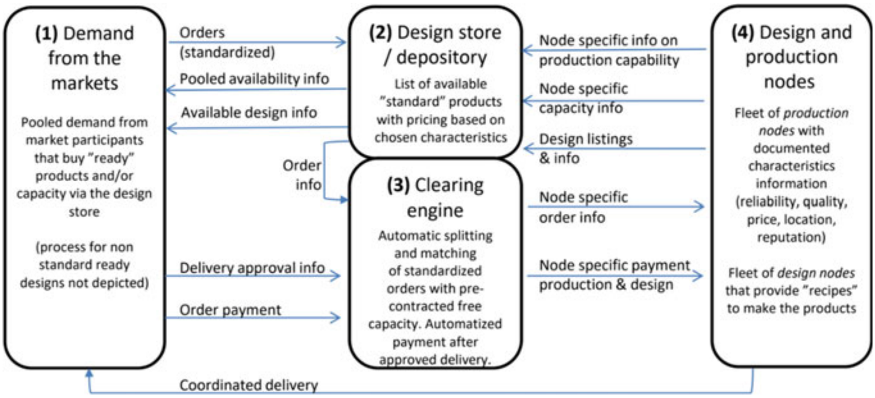


Fig. 2 Blueprint for a matching mechanism in a digital marketplace for the coordination of supply and demand for local manufacturing of standard products

and may include specification of speed at which the ready product must be delivered at the latest and must include the maximum acceptable price and may specify other requirements, for example the "reputation" required of the production nodes. Also, if there are requirements for "risk management" in terms of splitting large orders between multiple production nodes, such operations can be done either within the envisioned system or at the demand end. At some point, the credit of the demanding agent is checked (if not pre-approved).

The store is at the same time a depository of designs that are typically owned by the designer and that are used in the production of the products by the production nodes. Often the designers are owners of production nodes. Proprietary designs (for which the rights remain at the designer) can be added to the store by the design nodes and must include standardized information that is common and available for all designs. In a modern digital store-environment, products (here designs) will have a rating system that allows those who have bought the designs to rate them. The above-described design sharing constructions have been already implemented in the hobbyist 3D-printing online platforms [13]. In the case of the fully automated system proposed here, it must be able to distinguish between the design rating and the rating of the production node that has manufactured the design—both will have a separate rating of their own.

As the price of a given product is relative to each producer and the quantity purchased the pricing information is of high relevance in the store. Each production node must provide information about their node-specific capacity (quasi-real-time) and pricing (or pricing as a function of the quantity and delivery speed) for each design/product they are ready and able to manufacture to the store. The information is considered binding that is, non-negotiable and pre-approved by the production node; this is required for the market-mechanism to function properly. Also, the delivery information and cost for each production node must be considered ex-ante—this is now for the most part trivial, but if high quantities are demanded or demand is for delivery on short notice, the orders may have to be split between production nodes, causing a combined logistics-cost calculation to take place.

The orders (demand) and the supply are matched by the system (Clearing engine), as shown in Fig. 2, and if necessary, automatically split between the production nodes. After the matching of the order, the node-specific orders are created and matched with the necessary designs that are fetched from the store and relayed to the production nodes. In a highly automated configuration, the designs may even be automatically placed on a production queue of a modern additive manufacturing unit. Some issues with the proprietary designs remain, for example, how can one ensure that designs are not misused by the production nodes after they are received. Encryption of designs and on-demand automatic decryption of designs within the production devices or other ways to solve digital-rights issues in this context are interesting issues—for as long as good universal answers to these problems exist, trust remains a key issue within local manufacturing networks.

Payment of goods takes place according to a pre-approved schedule that may be based on triggers such as the delivery of goods (partly or fully), or on time-based tranches. The payment for the use of the designs may or may not follow the schedule for the payment of products. Payment can also be contingent on the delivery, in which case information on delivery will trigger payments. Pre-approved contracting and payment scheduling allows fully automated transactions to take place after triggering information is provided to the system.

3 Summary and conclusions

In the context of digital local manufacturing, data in various forms is the key commodity. The driver is the information about the demand that can be attracted if a well-functioning and reliable network for standardized product delivery is in place, in other words, when a credible marketplace for acquiring the demanded goods and services exists. In the modern world of business, the marketplace is typically digital. What we know of digital business in general holds also for the digital manufacturing business: smooth access to the marketplace, standardization of products and payment are key underlying factors for success, knowledge of the existence of the marketplace is pivotal for the demand to gather around it, and trust in many forms drives the business. The owner of the marketplace is the owner of the information and thus sits on a valuable commodity that can be used to create additional value in many ways.

If the evolution of digital marketplaces for 3D-printed products and additive manufacturing services will most likely concentrate at first around specific industry areas in a way that designs and products for the area are available via a given marketplace. The marketplaces may be driven by specific industry participants in collaboration or by third parties that even represent the demand. In a modern industrial environment, the demand for the envisioned system can, for example, be driven by a sufficiently large set of predictive maintenance driven facilities, where the demand is generated by the need to replace worn parts in the facilities—in such cases, the orders can realistically (and in line with the vision) be automated and handled directly by the predictive analytics systems [14]. In such a case the marketplace could, for example, be run by the maintenance-providing agents. A next step in the evolution of the marketplaces may be a similar type of consolidation that has been seen with established platforms, where various groups of "contexts" are represented on a single platform—sold items ranging from books to healthcare products, from collectibles to machine tools. The consolidation may be driven by platform technology or by excellent marketing and "making the platform known"—the platform that everyone knows will get the most demand traffic, which in turn will cause it to be most desirable from the point of view of the supply-side actors.

The platformization of digital manufacturing means that those that are early adopters of successful platforms may come out of the game as winners, also standards within winning platforms in terms of manufacturing technology and various other issues such as delivery and quality terms, may become overall de facto industry standards, which inevitably will also affect device manufacturers who are in no uncertain terms forced to adhere to standards or provide multi-standard devices. Digital interfacing will become more important than before—machines that are unable to communicate with the platform are losing out on business or must be converted via a "translator" to gain access—costs will be incurred in such cases. Pricing is driven down as happens in markets, where numerous competing actors are providing services—high-quality high-reliability service providers with good reputations, a connotation for trust, may reap higher added value.

The implications of the autonomous machine-to-machine digital information-based market mechanisms as discussed above are not known. What the demands of such mechanisms for contracting, cryptography, supply chain- and risk management are, is still an open question. What we know from past research helps to formulate relevant hypotheses and guides us forward, but what the most important specific challenges and drivers are remain unknown.

References

1. Strange, R., Zucchella, A.: Industry 4.0, global value chains and international business. Multinatl. Bus. Rev. **25**(3), 174–184 (2018). https://doi.org/10.1108/MBR-05-2017-0028
2. Li, L.: "China's manufacturing locus in 2025: with a comparison of 'made-in-China 2025' and 'industry 4.0.'" Technol. Forecast. Soc. Change **135**, 66–74 (2018). https://doi.org/10.1016/j.techfore.2017.05.028
3. Redelinghuys, A.J.H., Basson, A.H., Kruger, K.: A six-layer architecture for the digital twin: a manufacturing case study implementation. J. Intell. Manuf. (2019). https://doi.org/10.1007/s10845-019-01516-6
4. Negri, E., Fumagalli, L., Macchi, M.: A review of the roles of digital twin in CPS-based production systems. Procedia Manuf. **11**, 939–948 (2017). https://doi.org/10.1016/j.promfg.2017.07.198
5. Oztemel, E., Gursev, S.: Literature review of industry 4.0 and related technologies. J. Intell. Manuf. **31**(1), 127–182 (2020). https://doi.org/10.1007/s10845-018-1433-8

6. Lu, Y., Liu, C., Wang, K.I., Huang, H., Xu, X.: Digital twin-driven smart manufacturing : connotation, reference model, applications and research issues. Robot. Comput. Integr. Manuf. **61** (Apr 2019), 101837 (2020). https://doi.org/10.1016/j.rcim.2019.101837
7. Täuscher, K., Laudien, S.M.: Understanding platform business models: a mixed methods study of marketplaces. Eur. Manag. J. **36**(3), 319–29 (2018). https://doi.org/10.1016/j.emj.2017.06.005
8. Mentzer, J.T., DeWitt, T.W., Keebler, J.S., Min, S., Nix, N.W., Smith, C.D., Zach, Z.G.: Defining supply chain management. J. Bus. Logist. **22**(2), 2001 (2001)
9. Chen, H., Chiang, R., Storey, V.: Business intelligence and analytics: From big data to big impact. MIS Q. **36**(4), 1165–1188 (2012). https://doi.org/10.2307/41703503
10. Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D.: How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. Int. J. Prod. Econ. (2015). https://doi.org/10.1016/j.ijpe.2014.12.031
11. Trkman, P., Mc Cormack, K., de Oliveira, M.P.V., Ladeira, M.B.: The impact of business analytics on supply chain performance. Decis. Support Syst. **49**, 318–327 (2010)
12. Demirkan, H., Delen, D.: Leveraging the capabilities of service oriented decision support systems: putting analytics and big data in cloud. Decis. Support Syst. **55**(1), 412–421 (2013)
13. Rayna, T., Striukova, L.: From rapid prototyping to home fabrication: how 3D printing is changing business model innovation. Technol. Forecast. Soc. Chang. **102**, 214–224 (2016). https://doi.org/10.1016/j.techfore.2015.07.023
14. Urbani, M., Collan, M.: Additive manufacturing cases and a vision for a predictive analytics and additive manufacturing based maintenance model. In: Collan, M., Michelsen, K.-E. (eds.) Technical, Economic and Societal Effects of Manufacturing 4.0. Palgrave-McMillan (2020). https://doi.org/10.1007/978-3-030-46103-4_7

Digital Supply Chain and Blockchain: A Living Lab Perspective

Luca Mattia Gelsomino, Terrence Bos, and Michiel Steeman

Abstract Living labs are not new. However, they are often framed in literature as methods for user-centric innovation management. We argue that a living lab can be seen as a research methodology within the field of cooperative inquiry, as an operationalisation of concepts such as action research and engaged scholarship. To do so, we present the case of a living lab in the Netherlands, focused on data sharing technologies, and analyse how it generates research output with clear pragmatic value as well as addressing interesting research challenges. Finally, the paper positions the methodology against existing concepts within cooperative inquiry.

Keywords Living lab · Action research · Blockchain technology

1 Introduction

Living labs are not new. For several years now they have been the focus of attention for academics in the field of innovation management [e.g. 1]. Typically, living labs are positioned as methods to generate innovation in a collaborative and participatory design with the final user of said innovation. However, little attention has been paid to the possibility of living lab to be an effective methodology to generate research results. Within the scope of cooperative inquiry methods (e.g. action research, collaborative frameworks) a living lab can provide fertile grounds for research involvement in a practical problem and succeed in generating pragmatic value that would otherwise be difficult to reach. To elaborate on this point, we aim at focusing on the specific case of Blockchain Technology (BCT). BCT revolves around the creation of decentralized systems, self-executing contracts and intelligent assets controlled over the internet and it is forecasted to lead to an influx of decentralised control over supply chains. Combined with innovative technologies (e.g. internet of things) BCT

L. M. Gelsomino (✉) · T. Bos · M. Steeman
Windesheim University of Applied Sciences, Zwolle, The Netherlands
e-mail: lm.gelsomino@windesheim.nl

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_37

is expected to allow for improved information sharing, fostering trust and collaboration between companies, for the benefit of the entire supply chain. Several BCT-focused initiatives and consortia in the logistics and supply chain management areas have been launched in recent years, such as TradeLens, Batavia, we.trade and Marco Polo. Besides these initiatives, new and established technology providers (including TradeIX, Centrifuge, and Tradeshift), have started using BCT in the same areas.

Academic efforts are equally plentiful. For example, Treiblmaier [2] presents a framework for the development of BCT-based middle-range theories. Babich and Hilary [3] highlight academic challenges in the field of operations and supply chain management and focus on the need to regulate and govern the use of BCT within supply chains. Sternberg et al. [4] analyse BCT adoption for provenance of products, highlighting how attention towards the topic has so far translated in a scarcity of empirical evidence. Rogerson and Parry [5] recommends additional empirical evidence to be collected, with specific emphasis on longitudinal efforts.

In view of the difficulties that are presently experienced both practically (in effective adoption of an otherwise promising technology) and theoretically (in studying such adoption), we aim at provide evidence of how an alterative methodology, relying on the activities and results of a living lab focused on data sharing technology (including BCT) can provide insights on how to overcome existing obstacles and generate research results that are both of pragmatic value as well scientifically sound. Thus, our objective is to provide evidence of the relevance of living lab as research methodology, building on the practical case of a living lab in the Netherlands that aims at studying use case for data sharing in logistics and supply chain (including BCT adoption). By doing, we aim at expand the current view about existing theoretical perspectives on living lab [6] and reflect on the role of researchers in generating actionable and yet rigorous knowledge [7] within the domain of living labs, building on recent calls for leading and concurrent pathways from theoretical work to impact generation [8].

2 Literature Review

A Living lab is design research methodology to tackle complex problems. It aims at tackling the innovation process cooperatively, involving the innovation user. The living lab is traditionally identified by three dimensions: (i) being a test-bed for innovation involving final users, (ii) being a setting for users to define innovations and (iii) being an innovation network [1, 6]. A Living lab aims to foster innovation, but a living lab in itself can also be seen as an innovative approach, within the domain of process innovation [9]. This attains to the way in which innovation should be organized so the desired outcomes will be realized. Although living labs are not a novel methodology for design research, further research should be conducted to gather a better understanding of the design of the living labs related to the desired outcomes [6, 10]. A living lab can be seen as a form of design methodology which is research-led and has a participatory mindset; within that scope the interaction with users can be either open or closed and the focus can be on value capturing or value creation: a value capturing strategy focusses on exploiting the potentialities of the existing technology. A value creation strategy aims at exploring opportunities provided by new technology [1]. In terms of practical structure a living lab tends to be characterised as open or closed: in an open living lab everyone has the right to participate, while a closed strategy has a limited accessibility because users who participate are pre-selected.

Looking to living labs in more detail, Hossain et al. [10] distinguishes seven key characteristics of living labs: Real-life environments; stakeholders; activities; business models and networks; methods, tools and approaches; challenges and outcomes. According to Wolfert (2010) the living lab approach incorporates the strengths of the prototyping approach because of its iterative method and the aim for a rapid development of a workable solution. A disadvantage of this approach is that essential steps in the development process can be left out. In general terms, living lab literature sees those instruments as either environment for stakeholders (including users) to evaluate and experiment with innovation, a methodology for generating innovation or as systems aspiring to gather feedback from various contexts and innovation activities [10]. The link between living labs and cooperative inquiry, such as action research, is not entirely new in literature [11], and the impact that a living lab can have on research, especially around applied sciences, is noted [12]. However, to the best of our knowledge this has not yet been fully addressed.

3 Methodology

As anticipated, this study relies on the activities of Spark!, a living lab located in the Netherlands and focused on developing use cases around the topic of data sharing innovation in logistics and supply chain management, including BCT. The Spark! Living lab focuses on investigating and developing use cases in the field of data sharing applied to physical assets in supply chain and logistics. Beyond practically developing these use cases together with the companies involved, a main goal is to build a repository knowledge for parties, especially SME’s, that want to experiment with data sharing. The program started in 2019 will end in 2021. The living lab’s work revolves around ‘use cases’, practical issues related to the topics object of the lab that are experienced by one or more companies within a supply chain. The use cases are developed by a project board and supported by a validation board composed of both practitioners and researchers. The execution is delegated to the project manager and the subproject leaders. The subproject leaders are managing the use cases. For each use case an innovation approach is carried out in four steps: ideate, analyse, develop prototype and run experiment.

Within the context of Spark!, researchers perform activities through action research [13, 14]. The study specifically focuses on one large corporate (called ‘X’ in short) within Spark!, which joined the Lab to evaluate the potential for data sharing (including evaluating BCT adoption) improvement in their supply chain. Despite being part of a global group, the use cases object of the analyses were focused on the European branch of the company. The corporate's European branch focused on processing and freezing food collected mostly from European suppliers and distribute it to European customer in the horeca industry. The company is involved in two use cases, which were put forward by management and refined together with researchers from the lab. Data collection took place in several different forms: (i) an initial 'intake interview' was performed with every company involved in the use case, in which to map out the company supply chain, operations and financial processes, as well as measure key variables such as relationship and level of trust with other supply chain members; (ii) interview have been performed by the research team with other relevant parties that will be influenced by innovation streaming from the work of the living lab, such as a customer of the corporate; (iii) per 'use case', meetings have been held periodically to discuss and present the advancements in the work, including both the research team and representatives from the companies involved in the use cases; (iv) a survey sent to the X's suppliers to acquire core information about their view in respect to the use case that was of their interest; (v) several site visits at different companies, to validate information acquired through other methods, map processes and perform basic ethnographic research. Data collection was intertwined with feedback sessions, either in plenary consortium meetings or with practitioners, to provide feedbacks on the data collected and discuss next steps, actions planned and implementation.

4 Analysed Case

As discussed, the analysed case pertains to a food producer active in processing and freezing food for distribution to the horeca sector in Europe. Following the *'preliminary'* part of the cycle proposed by Maestrini et al. (2016), the research team first identified the appropriate areas for action. These were summarised in two use cases: on the supplier side the management of supplier certifications, and on the buyer side temperature management for condition goods.

4.1 Food Safety Certification

The first of the two analysed use cases pertains to certification for food safety. All direct suppliers of X are required (by current legislation) to provide a specific certificate attesting that they follow food safety requirements. The vast majority of suppliers are being certified by the same certification body. The certification process presents some clear challenges for the actors involved, namely: for X, an un-certified supplier cannot be included in the order scheduling. Considering that there appears to be some kind of seasonality in certification, X might find itself in the situation in which orders cannot be placed with several suppliers, since the due date for delivery falls after the expiration day of the certificates. For suppliers, a similar problem occurs. While on one side they might not be strongly incentivised to anticipate the renewal of a certificate (which cost money and results in certificate effectively lasting less than their nominal time span), they risk to loose orders from X. For the certification body, more transparency and visibility on the certification process will improve the quality of their services and provide an overall better experience to their direct customers (the supplier); it also means less time spent on managing request for additional information from X.

On top of this actor-specific challenges, all three parties lament that they have to spend time (and, thus, money) in calling, emailing and sharing information in unstructured manners, for example sharing Excel files with list of certified suppliers. This increases the probability of errors and the time required to properly manage the process. At the same time, it's unlikely that a simple online repository can serve the purpose of solving these issues. While the quality of the relationship between the parties cannot be described otherwise than very good, two issues are clear: (i) it remains a transactional principal-agent relationship and (ii) as the use case takes place within the European Union, the General Data Protection Reform (known as GDPR) applies. As such, privacy concerns are paramount: regarding point (i), there might be potential situations or cases in which a supplier might not want other parties to know the status of their certification process at a specific moment in time (for example due to issues that might cast doubts on their reliability); whether X accesses information about certification of a specific supplier (which likely include at least some data covered by GDPR), it should be with the explicit consent of that supplier.

More in general, a platform for data sharing of food certificates could be extended by all parties (X, certification body and suppliers) to multiple actors not involved in this specific use case, which makes the use of BCT more relevant and increase potential benefits. As such, the living lab focused on developing a prototype for a data sharing platform that allows the certification body to share food-safety certificates and other relevant information with X regarding the certification of X's suppliers, with the supplier's consent, with a certain degree of security (e.g. data immutability) provided by the adoption of BCT.

4.2 *Temperature Management for Conditioned Goods*

The second use case pertains to measuring and monitoring temperature from the end of the production process (where the product is frozen) and the arrival of the product to the warehouse of the logistics service provider, where it is then stored for delivery to customers. The current temperature management process is managed by physically 'sticking' temperature sensor in the boxes of products at pre-defined moments in time (either by X or one of the LSPs), the employee responsible for the measurement (e.g. the truck driver) records the temperature on paper and, later on, in an excel file. The Excel files with all the temperature are not per se shared with the other supply chain partners, unless there is a specific complaint from the downstream supply chain (i.e. the horeca sector). However, those are reported to be extremely rare. However, two points do make the use case relevant for X and the involved parties: (i) there are discussion between X and the two LSPs with respect to the temperature the moment in which the products are taken over by the LSP. The contrast comes from actors wanting both to minimise the energy (and, sometime, time) spent to freeze products. X aims at providing products at the highest possible temperature, while the two LSPs aims at taking over products that are already at the lowest possible temperature; (ii) X aims at improving and extending its distribution network, including for example modal transportation including inland sea and rail as well as road. This very likely implies additional legal entities involved, as well as more loading and unloading, which might affect temperature of the products and make temperature tracking more difficult.

As such, the living lab setup an experiment focused on installing sensors for real-time temperature monitoring throughout the entire process from the end of the production line to arrival at the LSP warehouse, shared with interested parties. Although the experiment is still ongoing, it will allow for higher control over temperature and identification of issues related to temperature management.

5 Results and Discussion

The high level of uncertainty that characterises the definition of the research problem tackled within the consortium affects both the research as well as the pragmatic results, and requires several iterations of the *'preliminary cycle'* to frame the problem correctly, meaning a constant reiteration between researcher and practitioners. This level of uncertainty is not limited to a potential lack of knowledge on a new phenomenon, as BCT can be, but is also a combination of other factors as, for example, the number of participants and the lack of visibility in the supply chain. It is a matter of *environmental uncertainty*.

A practical example comes from the second of the two use cases presented in the previous section. A first iteration with X led the research team towards the investigation of research possibilities within the temperature management from the LSP warehouse towards customers. Investigations with LSPs and customers led to the understanding that that process was not as critical as initially thought, and shifted the focus of the effort on the process between the end of production and the arrival at the LSP warehouse. Within this process, it was initially difficult to determine what the theoretical framework might be, meaning that the companies involved had a general 'hunch' that additional data could improve the process, but where not sure of what exactly could be done with additional data, or what specific mechanism could bring benefits. Thus a cycle of *'what can be done'* and *'what is useful'* was performed, to arrive to the definition of an experiment (in practical terms, not methodological) that could provide data and clarify to the involved parties the value of additional visibility on temperatures.

The second emerging perspective is related to the emerging of contradictions among actors in relation to the problem setting. Sometimes, parties simply have different perspectives that cannot be easily reconciled. For example, in the first of the two use cases there is a strong level of uncertainty regarding the behaviour of the other parties involved. After a first phase of relative alignment, X manifested some concern over a perceived lack of involvement from the side of the certification body, which they perceived as (potentially) opportunistic behaviour. A similar conflict revolves around involving suppliers: although they play a clear role in the use case, X and the certification body are both concerned about how and when to approach them. They are uncertain about their behavioural response. Even within X itself there are similar problems: suppliers relationship pass through key procurement employees that are, essentially, key account managers. Those stakeholders also present different perspectives than the managers involved in the living lab and are wary of opportunistic behaviour.

We see the presence of *behavioural uncertainty* as a limit to the pragmatic value of the research effort. In a more traditional approach (e.g. a case study) the detection of behavioural uncertainty could be, in itself, theoretically interesting and the focus of a potential contribution on, for example, understanding or explaining why BCT is adopted or not in a specific context. However, this would have limited pragmatic value, at least for the companies involved in the consortium. They are already aware that they are uncertain about other parties' behaviour, and don't need that to be explained to them. For them, the role of behavioural uncertainty in such pilots is clear and not difficult to predict, or understand. Therefore, behavioural uncertainty in this context becomes an obstacle to overcome to arrive to a problem definition that has pragmatic value for the companies involved. Arriving to a 'good' (meaning: both theoretically sound as well as pragmatically valuable) problem statement in this context implies an active effort from researchers to eliminate behavioural uncertainty, through direct involvement. In the specific case, feedback rounds and meetings with the different parties were specifically held, led by the research team, to work out differences in perspectives and limit the space for opportunistic behaviour, limiting behavioural uncertainty. This clarified the scope of the use case and led to identify a more pragmatically valuable research question, verting around understanding the role for digitalised data sharing on existing supply chain relationships.

5.1 Living Lab as a Research Method

The issues illustrated in the previous section lead to the strengths of the living lab approach from a research methodology perspective. These revolve around two main issues.

The first relates to how is it possible to define a theoretical framework leading to pragmatic value when high levels of uncertainty affect the area of investigation. Essentially: how to define a theoretical framework, select a research question and collect data to answer it, when there is no indication of where actionable impact (to refer again Voss, 2020) is going to be, because the practitioners are—while research takes place—working on developing innovation. From a researcher that does not take into consideration cooperative inquiry, in such a case there are two possible outcome: either to focus on what can be measured at the present moment (which, however, is likely of little pragmatic value, as explained above) or wait, as an external agent to the innovation process, until the context is ready to produce actionable results (as explained by Voss, 2020, p. 539), which however disincentives researchers. In this sense, the involvement of researchers in the living lab for the specific purpose of 'using' the living lab as a methodology for research activities allows to go beyond this level of uncertainty (through back and forth communication, room to experiment, and so on) and address pragmatic value.

This first point, however, could be extended to other forms of cooperative inquiry, and is not unique of this specific case under analysis; however, a second issue is indeed unique. This relates to ensuring validity and reliability of the data collected. When the issue under investigation presents high levels of complexity and its area of application spans naturally over multiple actors in a supply chain, it's natural to inquire with more scrutiny whether data collected is valid and reliable. First of all, a topic such as BCT adoption requires multiple perspectives, even on the research side, to be fully understood: supply chain management, computer science and general IT management knowledge are critical to ensure that data collected is both valid and reliable. The living lab helps in this sense by naturally including in the same team researchers from different domains. Second, as actors in the supply chain provide from time to time conflicting perspectives, it is important to have the possibility to constantly go back to them with new evidence and test new ideas. This seems to be technically more difficult in a typical action research environment, in which you should move from data gathering to analysis, to action, to implementation, to feedback [13]. What if the data collected is contradicting in nature and cannot lead to action? This is a concept expressed by Maestrini et al. [15], that illustrate how the 'traditional' action cycle needs additional steps to account for the 'messy' relationship between a buyer and a supplier, leading to the 'action cycle reloaded'. We argue that this concept should be extended when the relationships are even more than the single buyer–supplier dyad, and that the living lab is the right structure to do so, as it provides the possibility to naturally cycle back at essentially every step of the action research process.

We are not aiming at reinventing the wheel. Our positioning of the living lab is well within existing conceptualisation of cooperative inquiry. Our view of a living lab as research methodology clearly builds on existing literature on action research, which in turns is seen as 'the most dominating approach' [16] within engaged scholarship, the general idea of collaboration within a learning to enhance scientific knowledge as well as pragmatic value [17]. Moreover, in our view the living lab is also closed to what is known as a *'collaborative research framework'* between academics and practitioners [18], or more generally to the concept of generating actionable knowledge in the field of innovation [e.g. 7]. However, we argue that the living lab is a specific operationalisation of these concepts that differ from the one presented in those contribution in the sense that it provides restricted capacity (working on a selected number of use cases, while Gastaldi and Corso [2013, p. 76] mention working with 300 practitioners at the time) but at the same time focusing more on the specific use case, collecting in-depth data and allowing the process of building theory ‘from the particular to the general in small steps’, which is typical of action research [13].

In conclusion, what we argue is that a living lab is not necessarily, as depicted in literature so far, only a process of generating innovations in complex context, but that it has a role within cooperative inquiry methodologies for scientific research. It can, essentially, be used as a collaborative research framework to tackle specific research questions in areas in which there is a relatively high number of actors involved and high uncertainty in defining a theoretical framework. It allows to tackle uncertainty in the behaviour of supply chain partners and in the environment surrounding complex innovation processes through constant and repeated feedback between a multi-domain research team and practitioners involved in the process.

6 Conclusions

The main theoretical implication of our contribution lies in framing the living lab as a research methodology within cooperative inquiry. We provide some insights, based on the experience within one of those Labs, of the obstacles and insights that can emerge from adopting this perspective. In this sense, our practical implications relates to lesson learned in managing Spark!, which can provide useful insights for future endeavours of research teams that wishes to adopt a similar approach. The most relevant points can be summarised as followed: first, a critical issues arises in keeping stakeholders updated on the current situation of the projects. This is due to the volume and concurrency of the different activities going on at the same time within the Lab. The constant back and forth in one-on-one and consortium meetings that, on one side, allows for good data and good management, on the other side takes time and effort and, if not done properly, leads to disenfranchise of some stakeholders and limit the value of the overall effort. Second, the capacity of the lab is limited. In this sense, there is a clear trade-off between managing cases properly (also linking to the first point), so that the most valuable results (from both a practical and theoretical perspective) can be achieved, and extending the work to more use cases, which improve the chances for generalisability. Third, we made use of both graduate and undergraduate students in supporting and operative roles. This provides some relieve to the capacity issues, but on the other side leads to the involvement of students that, especially at the undergraduate level, have a tenure within the lab that is significantly shorter than the time required to reach most meaningful milestones. This implies that students are either involved in a research process that might not be self-contained or that they might have to step in mid-way through an existing process. Both of these lead to tension in managing the need for students to reach their own milestones (e.g. graduation) while at the same time provide useful work for the lab.

Finally, there are of course limitations. As the activities of Spark! are still undergoing, there's still unclarity on the final results that it can achieve, both practically and theoretically. This implies that our contribution at this stage cannot yet be fully supported by the empirical evidence coming from the activities of the lab. Moreover, even if Spark! would provide enough empirical evidence to support our claims, it would still be one single case, and our insights would have to be extended and validated in different settings.

Acknowledgements This paper is part of the research program Sustainable Living Labs, co-financed by the Dutch Research Council (NWO), the Ministry of Infrastructure and Water Management, Taskforce for Applied Research (SIA) and the Top Sector Logistics.

References

1. Dell'Era, C., Landoni, P.: Living lab: a methodology between user-centred design and participatory design. Creat. Innov. Manag. **23**, 137–154 (2014). https://doi.org/10.1111/caim.12061
2. Treiblmaier, H.: The impact of the blockchain on the supply chain: a theory-based research framework and a call for action. Supply Chain Manag. An Int. J. **23**, 545–559 (2018). https://doi.org/10.1108/SCM-01-2018-0029
3. Babich, V., Hilary, G.: Distributed ledgers and operations: What operations management researchers should know about blockchain technology. Manuf. Serv. Oper. Manag. **22**, 1–18 (2019). https://doi.org/10.1287/msom.2018.0752
4. Sternberg, H.S., Hofmann, E., Roeck, D.: The struggle is real: insights from a supply chain blockchain case. J. Bus. Logist. 1–17 (2020). https://doi.org/10.1111/jbl.12240
5. Rogerson, M., Parry, G.C.: Blockchain: case studies in food supply chain visibility. Supply Chain Manag. **25**, 601–614 (2020). https://doi.org/10.1108/SCM-08-2019-0300
6. Dekkers, R.: Perspectives on Living Labs as innovation networks. Int. J. Netw. Virtual Organ. **9**, 58 (2011). https://doi.org/10.1504/IJNVO.2011.040935
7. Martini, A., Gastaldi, L., Corso, M., Magnusson, M., Laugen, B.T.: Continuously innovating the study of continuous innovation: from actionable knowledge to universal theory in continuous innovation research. Int. J. Technol. Manag. **60**, 157 (2012). https://doi.org/10.1504/IJTM.2012.049439
8. Voss, C.: Towards an actionable and pragmatic view of impact. Acad. Manag. Discov. **6**, 538–539 (2020). https://doi.org/10.5465/amd.2019.0113
9. Kahn, K.B.: Understanding innovation. Bus. Horiz. **61**, 453–460 (2018). https://doi.org/10.1016/j.bushor.2018.01.011
10. Hossain, M., Leminen, S., Westerlund, M.: A systematic review of living lab literature. J. Clean. Prod. **213**, 976–988 (2019). https://doi.org/10.1016/j.jclepro.2018.12.257
11. Voytenko, Y., McCormick, K., Evans, J., Schliwa, G.: Urban living labs for sustainability and low carbon cities in Europe: towards a research agenda. J. Clean. Prod. **123**, 45–54 (2016). https://doi.org/10.1016/j.jclepro.2015.08.053
12. Claude, S., Ginestet, S., Bonhomme, M., Moulène, N., Escadeillas, G.: The Living Lab methodology for complex environments: Insights from the thermal refurbishment of a historical district in the city of Cahors, France. Energy Res. Soc. Sci. **32**, 121–130 (2017). https://doi.org/10.1016/j.erss.2017.01.018
13. Coughlan, P., Coghlan, D.: Action research for operations management. Int. J. Oper. Prod. Manag. **22**, 220–240 (2002). https://doi.org/10.1108/01443570210417515

14. Schoenherr, T., Rao Tummala, V.M., Harrison, T.P.: Assessing supply chain risks with the analytic hierarchy process: Providing decision support for the offshoring decision by a US manufacturing company. J. Purch. Supply Manag. **14**, 100–111 (2008). https://doi.org/10.1016/j.pursup.2008.01.008
15. Maestrini, V., Luzzini, D., Shani, A.B., Canterino, F.: The action research cycle reloaded: conducting action research across buyer-supplier relationships. J. Purch. Supply Manag. **22**, 289–298 (2016). https://doi.org/10.1016/j.pursup.2016.06.002
16. Simonsen, J.: A concern for engaged scholarship: the challenges for action research projects. Scand. J. Inf. Syst. **21**, 111–128 (2009)
17. Van de Ven, A.H.: Engaged scholarship: a guide for organizational and social research. (2007)
18. Gastaldi, L., Corso, M.: Academics as effective orchestrators of interorganizational change and development: the experience of Italian observatories. Res. Organ. Chang. Dev. **21**, 59–119 (2013). https://doi.org/10.1108/S0897-3016(2013)0000021006

Industrial Engineering and Operations Research

Proposal of a Firefly Algorithm with Three Types of Functionally Differentiated Fireflies

Masaki Nagai and Ikuo Arizono

Abstract Firefly algorithm (Firefly Algorithm; FA) is one of the swarm intelligence algorithms. FA is known as the algorithm that is based on the flashing behavior of fireflies and their associated behavioral patterns. Specifically, FA associates the objective function with the light intensity of fireflies, and searches for the optimal solution based on the feature that fireflies are attracted to those with higher light intensity than themselves. In this way, we can find the optimal solution. However, since FA only uses the information of the other firefly with light intensity stronger than itself, it may not be able to find the optimal solution if the firefly with the highest light intensity falls into the local solution. On the other hand, while fireflies in nature are naturally distinguished into males and females, in general FA, all fireflies are defined as the same functional group with no gender and uniform behavioral patterns. Therefore, FA do not have the diversity of behaviors and may not be possible to properly solve problems due to uniform behavior patterns. In this study, we propose an algorithm that distinguishes the sexes of fireflies in FA and defines different behavioral patterns for female fireflies. By achieving diversity in the movements of fireflies, we aim to avoid the defined FA system falling into local solution and find the optimal solution. The proposed FA system is applied to the maximization problem of mathematical function, and the effectiveness of the system is verified by comparing it with existing FA systems in previous studies.

Keywords Firefly algorithm · Optimization problem · Swarm intelligence

1 First Section

In recent years, solving optimization problems has become very important. On the other hand, many optimization problems are becoming larger and more complex every year, making it difficult to find an exact optimal solution [1]. Therefore, optimal solution algorithm called meta-heuristics, which can efficiently obtain approximate

M. Nagai · I. Arizono (✉)
Okayama University, Okayama, Japan
e-mail: pyvj6vab@s.okayama-u.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_38

solutions with a certain degree of good accuracy in a realistic time, are attracting attention [2]. Among them, swarm intelligence algorithm has been attracting attention as optimal solution algorithm known as ant colony optimization (ACO) and particle swarm optimization (PSO), which are inspired by the collective behavioral patterns of organisms [3–6]. Here, Firefly Algorithm (FA) exists as one of the swarm intelligence algorithms [7–9].

The Firefly Algorithm (FA), proposed by Yang [7] in 2007, is an optimization algorithm inspired by the flashing behavior of fireflies and related behavioral patterns. FA associates the objective function with the light intensity of fireflies, and searches for the optimal solution based on the feature that fireflies with higher light are attracted to it. In the conventional firefly algorithm, all fireflies have been defined as agents in the algorithm with the identical function and behavior pattern. However, since the function and behavior pattern of all fireflies are the same, FA do not have the diversity of behaviors and may not be possible to properly solve problems due to uniform behavior patterns.

In this study, we investigate a functionally differentiated firefly algorithm that considers the sex of the firefly. In addition, we define two types of functions in firefly groups, which are defined as female fireflies, after defining the functional differentiation by sex. In this way, we can achieve diversity in the behavioral patterns of fireflies in a functionally differentiated system. As the consequence, we verify that the diversity in the behavioral patterns of fireflies in the functionally differentiated system will give us better solutions.

2 Overview of the Yang's Firefly Algorithm (FA)

This section describes FA that will be compared in this study. FA is the algorithm for optimal solution search proposed by Yang [7], who was inspired by the flashing behavior of fireflies. A firefly as an agent in FA is attracted to other fireflies and searches for a global solution according to the following three rules:

- All fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex.
- Attractiveness is proportional to their brightness, thus for any two flashing fireflies, the less brighter one will move towards the brighter one. The attractiveness is proportional to the brightness and they both decrease as their distance increases. If there is no brighter one than a particular firefly, it will move randomly.
- The brightness of a firefly is affected or determined by the landscape of the objective function. For a maximization problem, the brightness can simply be proportional to the value of the objective function. Other forms of brightness can be defined in a similar way to the fitness function in genetic algorithm.

According to these rules, the individual fireflies in the FA system behave in such a way that the system as a firefly swarm searches for the optimal solution.

Here, the power to attract other fireflies is called attractiveness. The attractiveness of a firefly is determined by its l intensity, and the farther the distance between the firefly and the comparison target, the weaker the intensity of the light reaching the target. Therefore, the attractiveness of a firefly is expressed by the following equation:

$$\beta = \beta_0 e^{-\gamma r_{ij}^2}, \tag{1}$$

where γ is the light absorption coefficient, r_{ij} is the distance between fireflies i and j with position vectors x_i and x_j, and β_0 is the intensity of attraction when $r_{ij} = 0$. The distance r_{ij} is expressed by the following equation:

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d} \left(x_{i,k} - x_{j,k}\right)^2}. \tag{2}$$

Firefly i is attracted by firefly j, which has higher light than itself, and moves according to the following equation:

$$x_i = x_i + \Delta x_i, \tag{3}$$

$$\Delta x_i = \beta_0 e^{-\gamma r_{ij}^2}\left(x_j - x_i\right) + \alpha\left(rand - \frac{1}{2}\right). \tag{4}$$

Note that α is a coefficient that controls the random walk and is a constant in the conventional FA, and $rand$ is a uniform random number between 0 and 1. The fireflies with the highest light intensity move randomly without being attracted by other fireflies, and their movement is described by the following equation:

$$x_i = x_i + \alpha\left(rand - \frac{1}{2}\right). \tag{5}$$

3 Firefly Algorithm with Distinct Males and Females (FADMF) by Takeuchi et al.

Firefly Algorithm with Distinct Males and Females (FADMF) is one of the firefly algorithms proposed by Takeuchi et al. [10] in 2016. In the nature, the sex of fireflies is existent. Referred to this, in FADMF algorithm, the sex of fireflies as agents in algorithm is distinguished, and then the movements of female and male fireflies are defined separately. In nature, female fireflies have larger bodies and male fireflies have larger eyes, and male fireflies find female fireflies by using their larger eyes and move toward them. On the other hand, female fireflies are characterized by their low mobility.

Then, in FADMF, the function of male fireflies is defined in the same way as that of Yang's FA. On the other hand, the function of the female firefly is defined separately from that of the male firefly. Specifically, a distance attenuation factor V and an attractiveness attenuation factor W are introduced for the movement of female fireflies to reduce their movement distance in comparison with male fireflies and make them less attracted to distant fireflies. Also, it has been assumed that male fireflies are attracted to all fireflies but female fireflies are attracted only to male fireflies. The equation for the migration of female fireflies is defined by the following equation:

$$\beta = \beta_0 e^{-\gamma r_{ij}^2 / W}, \tag{6}$$

$$x_i = x_i + \Delta x_i / V. \tag{7}$$

Further, Takeuchi et al. [10] have given the coefficient α in Eq. (5) as the time-varying parameters for male and female fireflies, respectively. Then, the equation for the parameter α of the male firefly is given as follows:

$$\alpha(t) = \alpha(0)\left(10^{-4}/0.9\right)^{t/t_{\max}}. \tag{8}$$

On the other hand, the equation for the parameter α of the female firefly is given as follows:

$$\alpha(t) = \alpha(0)\left(10^{-4}/0.9\right)^{t/2t_{\max}}. \tag{9}$$

Here, $t_{\max}$ is the upper limit of the number of search loops

4 Proposal of a Firefly Algorithm with Three Types of Functionally Differentiated Fireflies

Although the behavior patterns defined separately for each sex in FADMF by Takeuchi et al. [10], these behavior patterns are defined by an identical behavior principle that each firefly is are attracted by more attractive fireflies. Therefore, the behavioral principle of FADMF is similar to FA. This means that FA and FADMF is weak in the system ability to avoid falling into a local solution. Then, the optimal solution may not be able to find by these algorithms.

In this study, we consider a new firefly algorithm due to firefly groups that follow several different behavioral principles. Due to the existence of firefly groups with several different behavioral principles, the proposed firefly algorithm will be diverse. In the proposed algorithm, there are three different functionally differentiated behavior patterns in agents. The first is the behavior pattern in a male agent, which is attracted to all fireflies with larger light than itself, just like the same functional agent in conventional FA [7] and the male agent in FADMF [10]. Then, the attraction strength, distance, and travel equations of the male firefly are the same as those of the FA and are expressed in Eqs. (1)–(4), respectively.

Further, the second and third are behavior patterns in two types of female agents. Note that these behavior patterns in two types of female agents in this paper are different from the behavior pattern of female agents in FADMF [10]. Although two types of behavior patterns in female fireflies in the proposed algorithm have a common feature in the point that they walk randomly, they have the respective behavior patterns each type of female fireflies. In concrete, there are two patterns of the female firefly behaviors of a long random distance travel and a short random distance walk. Remark that the behavior patterns of a long random distance travel and a short random distance walk are not fixed to individual female fireflies but are switched depending on the environment in which the female fireflies are placed. The magnitude of the random walk is set to be small for the top 50% of female fireflies in terms of light for all female fireflies, and the magnitude of the random walk is set to be large for the bottom 50% of female fireflies. Then, the random-walk of female fireflies with top 50% light intensity is described as

$$x_i = x_i + 0.5\alpha(t)\left(rand - \frac{1}{2}\right), \tag{10}$$

and the random walk of female fireflies with bottom 50% light intensity is described as

$$x_i = x_i + 5\alpha(t)\left(rand - \frac{1}{2}\right). \tag{11}$$

Remark that the firefly with the highest light intensity, whether it is even a male firefly or female firefly, moves randomly without being attracted by other fireflies, and the movement of the firefly with the highest intensity is described by Eq. (5) with the time-varying coefficient $\alpha(t)$ given by Eq. (8), where Then, the equation deriving the coefficient $\alpha(t)$ is described as Eq. (8) in all of the male and female fireflies.

In our proposed algorithm, the behavior principle is not identical. In concrete, the female firefly in the proposed algorithm behaves randomly based on her own situation and does not depend on the light intensity information of other fireflies. This feature clearly achieves functional differentiation from the conventional FA and FADMA. Through the functional differentiation by three behavior patterns in agents, we consider that the proposed algorithm can achieve system diversity and applicability. On the other hand, in FADMF algorithm by Takeuchi et al. [10], both behavior patterns are governed by a common behavioral principle in that they depend on the light intensity information of other fireflies.

5 Numerical Verification

In this section, the algorithms in the conventional FA, FADMF by Takeuchi et al. and the proposed algorithm are applied to the maximization problem of mathematical optimization and compared in terms of solving ability. The usefulness of the proposed algorithm is demonstrated by applying the three models to several multimodal functions and comparing the number of optimal solutions found and the maximum value found.

The bivariate function $f(x_1, x_2)$ as the testing function is represented by the following equation:

$$f(x_1, x_2) = 80 - \left[x_1^2 + x_2^2 - 10\cos(2\pi x_1) - 10\cos(2\pi x_2) + 20\right]. \tag{12}$$

The shape of the graph of $f(x_1, x_2)$ is shown in Fig. 1. The function $f(x_1, x_2)$ is a multimodal function and has a maximum value of 80 at $(x_1, x_2) = (0, 0)$. The fireflies are sorted in order of decreasing light intensity, and a loop is defined as the time until all the fireflies finish their migration and the light intensity is updated. A loop is defined as one trial until the number of loops reaches 500, which is 100 trials.

Next, we explain the discrimination method for finding the optimal solution. The light intensity of the firefly with the highest light intensity among 40 fireflies is recorded as the maximum value, and at each loop, the maximum value is compared with the solution found by the firefly. If the solution found by the firefly is larger, the maximum value is updated. This is continued until the number of loops reaches 500, and the maximum value at the end of 500 loops is recorded as the best solution. If the best value is within 0.05 error of the true maximum value 80 of the function,

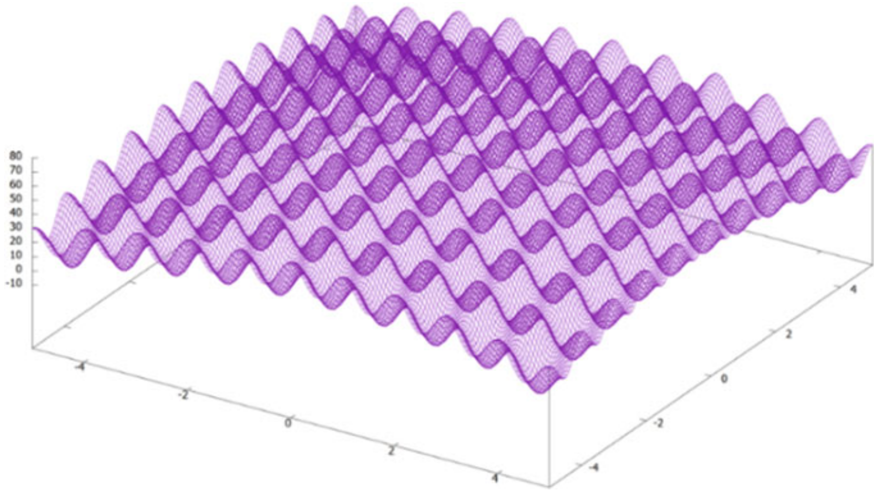


Fig. 1 The shape of $f(x_1, x_2)$

the best solution is found, and the number of optimal solution detections is counted. If no optimal solution is found, the search ability is evaluated by comparing the best values found. The average, maximum, and minimum values of the best values found are recorded, and it is verified whether it is possible to obtain a good value among them even when an optimal solution cannot be found.

6 Verification Results

As shown in Table 1, the proposed algorithm found the optimal solution more times than FA and FADMF, and it can be said that the proposed algorithm found the optimal solution as often as 98 at the end of the loop. Also, the proposed algorithm found the most optimal solution when applied to two bivariate functions different from function $f(x_1, x_2)$. Next, Table 2 shows that, the proposed algorithm could not find the optimal solution in some cases for function $f(x_1, x_2)$, the minimum value of the best solution found was 78.8987, which was higher than 75.0204 for FA and 75.0198 for FADMF. In addition, when applied to two bivariate functions different from the function $f(x_1, x_2)$, the proposed algorithm confirms that the minimum value of the best solution found is close to the maximum value of the function. In the case of a function with many local solutions, such as the function $f(x_1, x_2)$ shown in Eq. (12), FA and FADMF cannot escape from the local solutions, indicating that the optimal solution cannot be found.

Next, we explain Fig. 2. Figure 2 shows the time evolution of the best solution in a trial. In Fig. 2, the horizontal axis is the number of loops t and the vertical axis is the best solution.

In Fig. 2, for FA and FADMF, the value of the best solution suddenly becomes large at only one point, and then there is no change in the value of the best solution, indicating that the two algorithms have fallen into a local solution. On the other hand, in the proposed algorithm, the value of the best solution changed significantly at two

Table 1 Number of times an optimal solution was found in 100 trials in $f(x_1, x_2)$

Algorithm	100 loops	300 loops	500 loops
FA	5	7	8
FADMF	11	12	12
Proposed algorithm	58	84	98

Table 2 Average, maximum, and minimum of the best values for 100 trials in $f(x_1, x_2)$

Algorithm	Average	Max	min
FA	78.2293	79.9999	75.0204
FADMF	79.1612	79.9999	75.0198
Proposed algorithm	79.9014	79.9999	78.8987

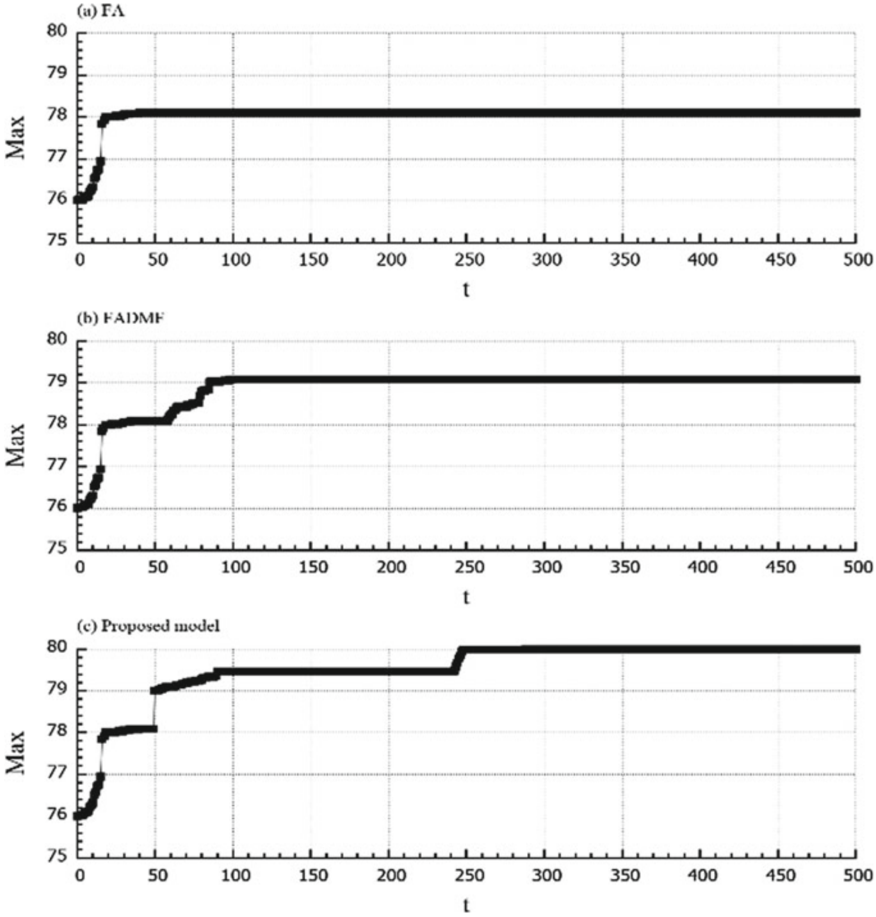


Fig. 2 Time evolution of the best solution in a trial

points, and then our algorithm was able to search for a better solution even after some time passed.

7 Conclusion

In the proposed algorithm, by having a female firefly that randomly walks on the system, the female firefly randomly searches for a region with a good solution even when the male firefly falls into a local solution, and the solution of the system as a whole can escape from the local solution. By differentiating female fireflies according to the size of their random walks, female fireflies in relatively good solutions do not move significantly from that state and maintain their good solution state. On the other hand, female fireflies in relatively unfavorable solution states are set to move a larger distance, so that these fireflies can move a larger distance to reach the region that gives them better solution states. In this study, the proposed algorithm with such functional differentiation enabled us to construct an algorithm with a better solution seeking ability than the conventional FA or Takeuchi et al.'s FADMF.

References

1. Michael, D.I.: Optimization and economic theory michael. SIAM Soc. Ind. Appl. Math. (1971)
2. Yang, X.-S.: Nature-inspired metaheuristic algorithms. Luniver Press (2010)
3. Dorigo, M.: Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Trans. Evol. Comput. **1**(1), 53–66 (1997)
4. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm intelligence. Natural to Artificial Systems, Oxford University Press (1999)
5. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proceeding of IEEE International Conference on Neural Networks, pp. 1942–1948. Piscataway, NJ (1995)
6. Krishnanand, K.N., Ghose, D.: Theoretical foundations for rendezvous of glowworm-inspired agent swarms at multiple locations. Robot. Auton. Syst. **56**(7), 549–569 (2008)
7. Yang, X.-S.: Firefly algorithms for multimodal optimization. Stochast. Algorithms Found. Appl. 169–178 (2009)
8. Shindo, T., Xiao, J., Kurihara, T., Morita, K., Jin'no, K.: Analysis of the dynamic characteristics of firefly algorithm. In: Proceeding of the of IEEE CEC2015, pp. 2647–2652 (2015)
9. Johari, N.F., Zain, A.M., Noorfa, M.H., Udin, A.: Firefly algorithm for optimization problem. Appl. Mech. Mater. **421**, pp 512–517 (2013)
10. Takeuchi, M., Matsushita, H., Uwate, Y., Nishio, Y.: Investigation of optimal ratio of males to females in firefly algorithm. J. Signal Process. **20**(4), 153–156 (2016)

Applying Overall Equipment Effectiveness to Pipe Jacking Tunneling in Construction Industry

Horng-Chyi Horng and Wen-Zhuan Xu

Abstract Overall Equipment Effectiveness (OEE) is a performance evaluation tool that measures equipment productivity with key indicators. In addition to construction industry, OEE has been widely used in different industries, such as manufacturing, food processing, semiconductor, medical service, etc. However, the characteristics of construction industry are the same as that of most manufacturing, that is, profits and revenue come from the effectiveness of machinery and equipment usage. Therefore, this research develops a method called equipment performance management system (EPMS) to evaluate the performance of the propulsion equipment of pipe jacking tunneling using OEE. This EPMS is applied to a real world case study to demonstrate how to define, calculate, and analyze these performance indicators of the pipe jacking tunneling machinery and equipment. The case study proves the feasibility of the research method, which is helpful for reflecting operating performance and identifying root causes of potential problems in time so that immediate improvement or rectification actions can be implemented to avoid further losses of the company.

Keywords Overall equipment effectiveness (OEE) · Construction engineering · Underground pipeline engineering · Pipe jacking tunneling machine

1 Introduction

Nakajima [1, 2] introduced a method called Overall Equipment Effectiveness (OEE) under the concept of Total Production Maintenance (TPM) to evaluate business performance. OEE is a performance evaluation tool that can measure equipment productivity with key indicators. It reflects production losses by calculating availability, performance rate, and quality rate, and provides targets for continuous

H.-C. Horng (✉)
Chaoyang University of Technology, Taichung, Taiwan
e-mail: hchorng@cyut.edu.tw

W.-Z. Xu
Dejia Tech Engineering Co., Ltd., Taichung, Taiwan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_39

improvement to enhance equipment productivity. In addition to manufacturing, OEE is also widely used in different industries, such as food processing industry, semiconductor industry, and medical service industry.

The profits and revenue of construction industry are the same as that of most manufacturing, that is, come from the effectiveness of machinery and equipment usage. Moreover, machinery and equipment are especially important to the construction engineering industry than the manufacturing industry. The suspension of construction operations due to machinery and equipment problems will not only cause delays in the construction project, inconvenience for passers-by and residents, but also affect the image of the company and induce intangible losses. Therefore, equipment performance evaluation has particularity and potential feasibility in the construction engineering industry. Therefore, this research selects the construction engineering industry's underground pipeline engineering and the pipe jacking tunneling machinery and equipment as the research object. It is expected to provide companies with a systematic approach to promote equipment performance evaluation, which will be applied to the diagnosis of equipment efficiency and provide directions for companies to improve their overall performance.

1.1 OEE

OEE is a quantitative indicator to measure equipment productivity, which can be used to express the ratio of actual production capacity to target production capacity. In other words, OEE is an important basis for evaluating whether equipment assets have achieved return on investment, and can also be used as a way to evaluate the effectiveness of production system improvement tool. The OEE indicator is composed of three major elements, as in Eq. (1), namely 1. Availability (A), the percentage of time the equipment actually operates, to evaluate the available time of the equipment in the current state; 2. Performance rate (P), the production speed of the equipment; 3. Yield rate or quality rate (Q), that is, the good product ratio of equipment production [3].

$$OEE = A \times P \times Q \tag{1}$$

Availability (A). The availability is the time utilization of the equipment, and can be expressed as the ratio of the actual operating time to the planned operating time as in Eq. (2) which specifically reflects the stop loss of equipment failures and adjustments. The planned operating time is the total shift time minus planned maintenance of the equipment. The planned operating time deduct the stop losses, the time during which the equipment is not in operation, such as unplanned maintenance, fault repair, setup and changeover, and equipment idleness, is called the actual operating time.

$$Availability\,(A) = actual\,operating\,time/planned\,operating\,time \tag{2}$$

Performance rate (P). The performance rate is composed of operating speed rate and net operating rate, as shown in Eq. (3). The performance rate is used to compare the difference between the design speed and the actual manufacturing speed of the equipment, without considering the production quality. The operating speed rate can reflect the difference between the designed standard cycle time and the actual cycle time of the equipment. If the equipment production is not disturbed and the production speed (actual cycle time) is also stable, the expected output of the equipment will be the ratio of the operating hours to the actual cycle time. But during equipment operation, there may be minor disturbances or short-term shutdowns, causing the actual production speed to be different from the designed production speed of equipment. The net operating time is the actual output in the working hours multiplied by the actual production cycle time of the product.

$$\begin{aligned} Performance\,rate\,(P) &= operating\,speed\,rate\; \times net\,operating\,rate \\ &= net\,operating\,time/actual\,operating\,time \end{aligned} \tag{3}$$

Quality rate (Q). The quality rate refers to the ratio of good products after deducting defective products from the actual output during the actual operating time. The denominator is the actual output, and the numerator is the actual output minus the number of defects and the number of reworks. It can also be viewed as the value time divided by the net operating time, as shown in Eq. (4). The value time refers to the yield of good products expressed in terms of time, that is, the number of good products multiplied by the standard cycle time of the equipment.

$$\begin{aligned} Quality\,rate\,(Q) &= (actual\,output - number\,of\,defects)/actual\,output \\ &= value\,time/net\,operating\,time \end{aligned} \tag{4}$$

Substituting Eqs. (2), (3), and (4) into Eq. (1), OEE can be expressed as the value time divided by the planned operating time, as shown in Eq. (5).

$$OEE = value\,time/planned\,operating\,time \tag{5}$$

OEE is a very popular tool for evaluating the productivity in manufacturing processes [4–6], maintenance [7–9], supply chain [10, 11], and enterprise resources [12, 13]. In addition to manufacturing industry, it has also been implemented in other industry or field, such as mining [14], transportation [15, 16], food industry [17, 18], and sustainability [19, 20].

1.2 Productivity in Construction Industry

Productivity is among the top three project performance indicators addressed over the last decade in ASCE's Journal of Construction Engineering and Management

[21]. Braglia et al. [22] reviewed a total of 25 articles in the field of indicators for construction productivity and found that many of them are cost-based approaches such as Tomas et al. [23] and Ayele and Fayek [24], where costs incurred are the only input factors considered as input variables in their productivity metrics. The main drawback of cost-based approaches is that cost induced by low productivity of the construction project cannot be found and intervened in time to minimize losses. Other reviewed articles address construction productivity, on the other hand, by placing emphasis on country-specific measurement, work-specific or project-specific applications, or aggregation to industry level.

Braglia et al. [22] further developed a metric called Overall Construction Productivity (OCP) based on the framework of OEE and lean manufacturing concepts to identify losses and causes of Engineering-to-Order (ETO) supply chain of a hospital construction project. Weerasinghe and Silva [25] directly applied the concept of OEE to qualifying machine effectiveness in building industry. However, as far as the development and research directions in pipe jacking and micro-tunneling field [26], there is a lack of research on evaluating the performance of the propulsion equipment of pipe jacking tunneling using OEE.

2 Methodology

The research method of this study takes four steps: 1. establishing the operation process of equipment performance management system (EPMS); 2. analyzing the loss structure of equipment; 3. defining and symbolling each equipment performance indicator (EPI), along with its calculation method; and 4. Constructing the analysis procedure of EPI.

2.1 The Operation Process of EPMS

To effectively apply OEE to evaluate the performance of the propulsion equipment of pipe jacking tunneling, the operation process of EPMS must be established first. As in Fig. 1, operating data collection is switched on once the propulsion equipment of pipe jacking tunneling was delivered to the construction site. The office of engineering management then calculates and monitors EPIs in real time, and also generates EPI reports periodically. These reports are then submitted to a periodic executive meeting to review all EPI. If any EPI are abnormal, a problem-solving team will analyze immediately to find out root-causes as well as implement corrective measures. Any corrective measure is evaluated by an improvement objective to see whether standardization is necessary to be made for guiding future operation of the propulsion equipment.

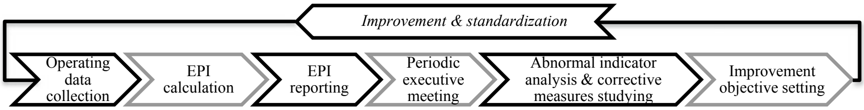


Fig. 1 The operation process of equipment performance management system (EPMS)

2.2 *The Loss Structure of Propulsion Equipment*

As defined in Sect. 1 for OEE and illustrated on the right-side diagram of Fig. 2, the planned operating time is the total operating time minus planned downtime, such as maintenance and downtime caused by improper schedule of the equipment. The planned operating time deduct the stop losses, the time during which the equipment is not in operation, such as unplanned maintenance, fault repair, setup and changeover, and equipment idleness, is called the actual operating time. During equipment operation, there may be minor disturbances or short-term shutdowns, causing the actual production speed to be different from the designed production speed of equipment (so-called speed loses). The net operating time is the actual output in the working hours multiplied by the actual production cycle time of the product. Finally, the value time refers to the yield of good products expressed in terms of time, that is, the number of good products multiplied by the standard cycle time of the equipment. Therefore, quality losses are number of defects multiplied by the standard cycle time of the equipment.

During the advancement of the propulsion process, there may be irregular advancing or re-advancing due to cement pipe rupture, causing the actual advancing length to be longer than the original designed advance length. These extra advancing lengths are the defects in propulsion. The quality losses are thus the extra equipment operating time to complete these extra lengths. The extra equipment operating time causes by quality problems cannot be effectively separated from extra equipment time caused by speed losses during the propulsion process. Therefore, as showed in Fig. 2, the speed loses and quality losses are combined into speed & quality loses (SQL) on the right-side diagram.

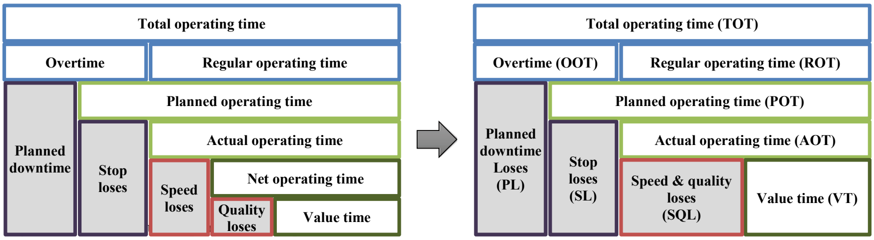


Fig. 2 The loss structure of propulsion equipment

Table 1 Major EPIs for the performance of the propulsion equipment

EPI	Symbol	Calculation
Propulsion progression proportion	**PPP**	*Advancing length completed / Total construction length*
Utilization	**U**	*POT/TOT*
Overtime ratio	**OT**	*OOT/ROT*
Availability	**A**	*AOT/POT*
Performance rate (revised)	**PQ**	*VT/AOT*

2.3 The Definitions of EPIs

In addition to OEE, there are other EPIs necessary for reflecting operating performance and identifying root causes of potential problems in time. Table 1 lists these EPIs, as well as their symbols and calculations.

2.4 The Analysis Procedure of EPI

The analysis procedure of EPI and corresponding corrective actions for each poor EPI performance are illustrated in Fig. 3. A total of four stages in the analysis procedure: 1. Making sure propulsion progression proportion (PPP) is ahead of project schedule; 2. Making sure utilization (U) is reasonable when taking into account overtime ratio (OT); 3. Making sure revised performance rate (PQ) is good enough; and 4. Making sure acceptable availability (A) is achieved.

3 Real-World Case Study

Case study of the EPMS utilizes company AAA, a construction engineering company in Taiwan who undertakes government sewage engineering bids and business related to propulsion equipment maintenance/repairs. During research period, company AAA has a sewage engineering project that utilization two set of propulsion equipment. The following demonstration of the analysis procedure of EPIs uses data in June, 2020. At review point, PPP of the project is 4.55% higher than original scheduled.

In the second stage of analysis procedure, confirming U is reasonable when taking into account OT. The diagram on the left-side of Fig. 4 shows the proportion of the case study divided into ten-week period of the propulsion equipment time allocated to the construction site. When PPP is reached, the case study's U is high, which means that the remaining capacity of the equipment is low. The analysis can be directed to the appropriateness of OT.

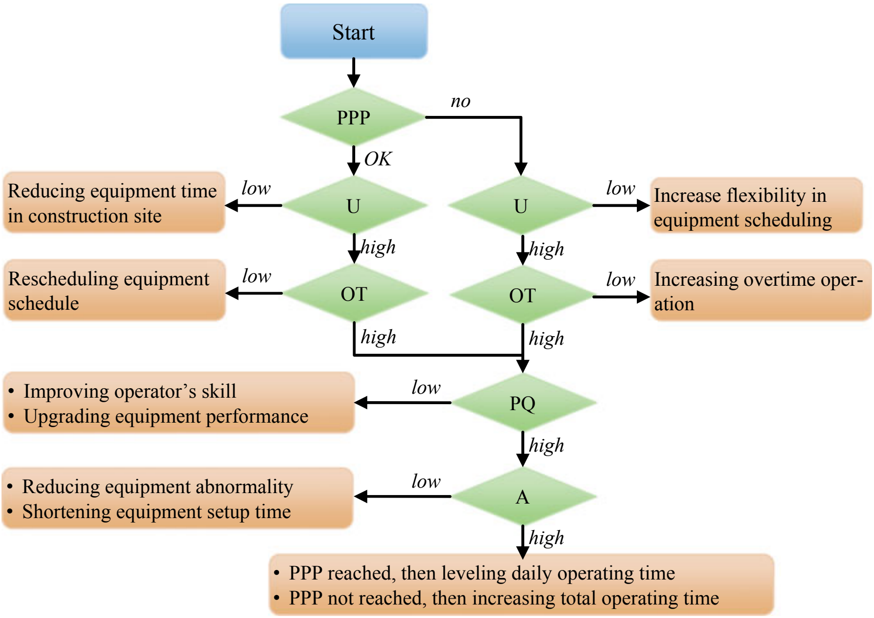


Fig. 3 The analysis procedure of EPI and corrective actions

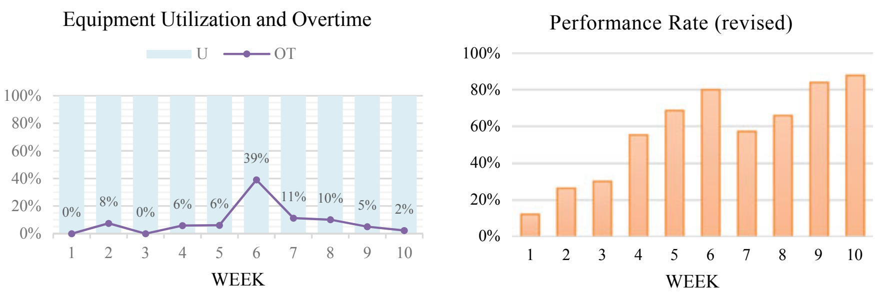


Fig. 4 The U and PQ of the case study

The third stage is to understand the degree of PQ. The diagram of the right-side of Fig. 4 shows the propulsion equipment PQ of Company AAA in ten-week period. It can be observed that the highest is 87.83% in the tenth week. PQ below 50% occurred in the first, second, and third week. Evidently the degree of PQ is poor. The direction of improvement in this situation is to conduct education and training, improve the technical ability of the propulsion operator, or overhaul and maintain the equipment to improve the performance of the propulsion equipment.

The fourth stage is to evaluate the extent to which the propulsion operation is disturbed and the failure to operate smoothly, that is A. During the propulsion operation, the amount of time lost during the stop will directly affect A. The low value of

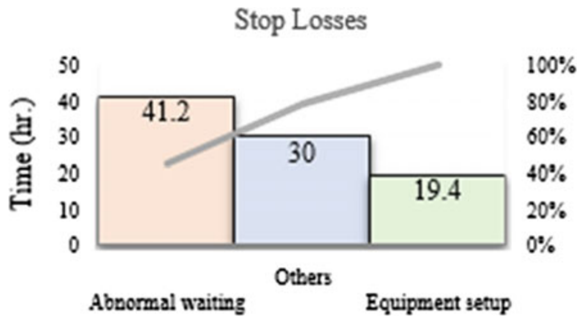


Fig. 5 The major stop losses of the case study

A means that the propulsion operation often suffers from equipment abnormalities or propulsion preparation time and other stop losses. Figure 5 is a Plato plot that summarized the stop losses in three categories: equipment setup, abnormal waiting, and other losses. Apparently, the most urgent corrective actions company AAA has to do are avoiding equipment abnormalities and shortening equipment setup time.

4 Conclusions

There are two major factors that affect the overall competitiveness of the construction engineering project. One factor is safety, quality, and completion on schedule. The other is resources such as manpower, equipment, and materials are used effectively. Safety, quality, and delivery time can increase market competitiveness, whereas the effectively usage of resources can reduce costs and losses, and thus increase overall profits. This research develops a method called EPMS to evaluate the performance of the propulsion equipment of pipe jacking tunneling using OEE. This EPMS is applied to a real world case study to demonstrate how to define, calculate, and analyze these performance indicators of the pipe jacking tunneling machinery and equipment. The case study proves the feasibility of the research method, which is helpful for reflecting operating performance and identifying root causes of potential problems in time so that immediate improvement or rectification actions can be implemented to avoid further losses of the company.

References

1. Nakajima, S.: Introduction to Total Productive Maintenance (TPM). Productivity Press MA Cambridge (1988)
2. Nakajima, S.: TPM Development Program. Productivity Press, MA Cambridge (1989)
3. Dal, B., Tugwell, P., Greatbanks, R.: Overall equipment effectiveness as a measure of operational improvement—a prectical analysis. Int. J. Oper. Prod. Manage. **20**(12), 1488–1502 (2000). https://doi.org/10.1108/01443570010355750

4. Huang, S.H., Dismukes, J.P., Shi, J., Su, Q., Razzak, M.A., Bodhale, R., Robinson, D.E.: Manufacturing productivity improvement using effectiveness metrics and simulation analysis. Int. J. Prod. Res. **41**, 513–527 (2003)
5. Anand, R.: Cloud computing OEE for reducing production downtime. SAE Int. J. Mater. Manuf. **6**(3), 481–486 (2013)
6. Hwang, G., Lee, J., Park, J., Chang, T.W.: Developing performance measurement system for internet of things and smart factory environment. Int. J. Prod. Res. **55**, 2590–2602 (2017)
7. Blanchard, B.S.: An enhanced approach for implementing total productive maintenance in the manufacturing environment. J. Qual. Maint. Eng. **3**(2), 69–80 (1997)
8. Shetty, D., Ali, A., Chapdelaine, J.J.: A model for the total productive manufacturing assessment and implementation. J. Adv. Manuf. Syst. **8**(2), 117–136 (2009)
9. Gupta, P., Vardhan, S.: Optimizing OEE, productivity and production cost for improving sales volume in an automobile industry through TPM: a case study. Int. J. Prod. Res. **54**, 2976–2988 (2016)
10. Buchmeister, B., Friscic, D., Lalic, B., Palcic, I.: Analysis of a three-stage supply chain with level constraints. Int. J. Simul. Modell. **11**(4), 196–210 (2012)
11. Perez Loaiza, R.E., Olivares-Benitez, E., Miranda Gonzalez, P.A., Guerrero Campanur, A., Martinez Flores, J.L.: Supply chain network design with efficiency, location, and inventory policy using a multiobjective evolutionary algorithm. Int. Trans. Oper. Res. **24**(1–2), 251–275 (2017)
12. Zammori, F., Braglia, M., Frosolini, M.: Stochastic overall equipment effectiveness. Int. J. Prod. Res. **49**, 6469–6490 (2011)
13. Golda, G., Kampa, A., Paprocka, I.: Analysis of human operators and industrial robots performance and reliability. Manage. Prod. Eng. Rev. **9**(1), 24–33 (2018)
14. Elevli, S., Evli, B. E.: Performance measurement of mining equipments by utilizing OEE. Acta Montan. Slovaca Ročník **15**, 95–101 (2010)
15. Garcia-Arca, J., Prado-Proda, J.C., Fernández-González, A.J.: Integrating KPIs for improving efficiency in road transport. Int. J. Phys. Distrib. Logist. Manag. **48**(9), 931–951 (2018)
16. Pinto, M.M.O., Goldberg, D.J.K., Cardoso, J.S.L.: Benchmarking operational efficiency of port terminals using the OEE indicator. Marit. Econ. Logistics **19**, 504–517 (2017)
17. Tsarouhas, P.H.: Evaluation of overall equipment effectiveness in the beverage industry: a case study. Int. J. Prod. Res. **51**, 515–523 (2013)
18. Tsarouhas, P.H.: Improving operation of the croissant production line through overall equipment effectiveness (OEE): a case study. Int. J. Product. Perform. Manag. **68**(1), 88–108 (2019)
19. Domingo, R., Aguado, S.: Overall environmental equipment effectiveness as a metric of a lean and green manufacturing system. Sustainability **7**(7), 9031–9047 (2015)
20. Durán, O., Capaldo, A., Durán Acevedo, P.: Sustainable overall throughputability effectiveness (S.O.T.E.) as a metric for production systems. Sustainability **10**(2), 362 (2018)
21. Jin, R., Zuo, J., Hong, J.: Scientometric review of articles published in ASCE's journal of construction engineering and management from 2000 to 2018. J. Constr. Eng. Manag. **145**(8), 06019001 (2019)
22. Braglia, M., Dallasega, P., Marrazzini, L.: Overall construction productivity: a new lean metric to identify construction losses and analyse their causes in engineer-to-order construction supply chains. Production Plann. Control 1–18 (2020)
23. Thomas, H.R., Maloney, W.F., Horner, R.M.W., Smith, G.R., Handa, V.K., Sanders, S.R.: Modeling construction labour productivity. J. Constr. Eng. Manag. **116**(4), 705–726 (1990)
24. Ayele, S., Fayek, A.R.: A framework for total productivity measurement of industrial construction projects. Can. J. Civ. Eng. **46**(3), 195–206 (2019)
25. Weerasinghe, H.M., De Silva, D.H.: Quantification of machine effectiveness in building and construction industry. In: 2017 National Information Technology Conference, NITC 2017, 13–15 (2017)
26. Sterling, R.L.: Developments and research directions in pipe jacking and microtunneling. Under. Space **5**(2020), 1–19 (2018). https://doi.org/10.1016/j.undsp.2018.09.001

How Moderator Variables Affect Scheduling Objectives in Unpaced Mixed-Model Assembly Lines

Frederik Ostermeier, Nikolai West, and Jochen Deuse

Abstract Besides the sequence itself also additional factors serving as moderator variables affect the value of scheduling objectives. For mixed-model assembly lines, especially number and heterogeneity of different products, their volume mix proportions, average workload of the jobs to process and the degree of grouping of identical jobs within the sequence play a major role. By means of a simulation study based on data from a real unpaced mixed-model assembly line in the automotive industry, this work analyzes the impact of these moderating variables on various scheduling objectives. The analyzed scheduling objectives encompass flow-related objectives like mean flow time, productivity-related objectives like makespan, customer-related objectives like mean earliness, the supplier-related objective part usage rate variation and the human-related objectives mean learning effect and mean deterioration effect per job. Simulation scenarios are defined that differ regarding number and heterogeneity of products from three homogeneous to seven more heterogeneous products. Within every simulation scenario the volume mix proportions of the products, and inherently also the average workload of jobs, are systematically varied. Every simulation scenario is analyzed for five sequence types differing in the degree of grouping of identical jobs. For almost all scheduling objectives, strong dependencies on the volume mix proportions can be perceived, particularly for mean flow time. Homogeneous volume mixes with a dominating product in the mix often lead to other objective values compared to heterogeneous volume mixes that allow using alternation effects between various products in a sequence. Concerning degree of grouping, while some scheduling objectives like part usage rate variation are always strongly affected by the degree of grouping for every volume mix, other objectives like throughput show strong dependence only for some mixes and makespan

F. Ostermeier (✉) · N. West · J. Deuse
Institute for Production Systems, Technical University of Dortmund, Dortmund, Germany
e-mail: frederik.ostermeier@bmw.de

F. Ostermeier
BMW Group, Munich, Germany

J. Deuse
Centre for Advanced Manufacturing, University of Technology Sydney, Ultimo, Australia

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_40

does not even show any tendency. Average workload plays a less important but still recognizable role as moderator variable for most objectives except for throughput for which workload is a major explanatory factor. Number and heterogeneity of different products has a strong impact on mean learning and deterioration effects.

Keywords Mixed-model assembly lines · Flowshop scheduling · Moderator variables

1 Introduction

The impact of a schedule as independent variable on a scheduling objective as dependent variable does not only depend on the schedule itself but also on several third variables, the moderator variables. A moderator variable is a "variable that affects the direction and/or strength of the relation between an independent [...] and a dependent [...] variable" [1]. Since the early days of scheduling, it has been known that these variables affect the height of the objective values and the desirability of schedules [2]. There is a plenitude of moderator variables with impact on scheduling, especially on mixed-model assembly line scheduling.

On a mixed-model assembly line, several similar products, that can be understood as variants, are assembled in a non-restricted sequence of jobs [3]. These variants are according to [4] products whose characteristics vary from a basic product configuration only to an extent that still allows the joint production of these products on the same line due to a sufficient production-related similarity in operations, and even in a sequence with no restriction on the assignment of jobs to a position as no setups are required between units of different variants.

Scheduling a mixed-model assembly line is challenging as the impacts of a schedule on the scheduling objectives are not always intuitive and straight-forward. Tactical design decisions resulting in the line balance often play a major role in moderating the impacts of a schedule since the schedule interacts with the line balance by mitigating imbalances between jobs of different variants and between stations where a job has different processing times [5]. If a line is almost perfectly balanced, scheduling would play merely a minor role [6]. Indeed, industrial mixed-model assembly lines are rarely perfectly balanced [7] and the mitigation effect of scheduling is crucial.

These interactions between the schedule and the line balance have yet been addressed by various studies (cf. [8]). Less academic interest has been set on the operational moderator variables. While the line balance is a static input, the operational moderator variables may change on a daily basis [9, 10]. These encompass, among others, number and heterogeneity of different products, their volume mix proportions, average workload of the jobs to process and the degree of grouping of identical jobs within the sequence. This work will put its focus on the impact of these operational moderator variables on the relation between the schedule and scheduling objectives. What is known in literature about their impact, is briefly outlined in the following:

- **Product mix**: The product mix denotes the number and heterogeneity of products for instance regarding required operations and the resulting processing times on stations. Mixed-model assembly lines generally perform worse if the product mix becomes too heterogeneous and too many different products are produced [11].
- **Volume mix**: Strongly related with the product mix is the volume mix, i.e. the distribution of volume percentages per product, influencing the performance of mixed-model assembly lines significantly [12, 13]. For instance, [10] showed the effect that the volume mix has on throughput in a production system with multiple products.
- **Workload**: Average workload, denoting the sum of processing times of the jobs currently in the production system, has long been known for its moderating impact on scheduling [2, 14]. Moreover, the workload does not only influence the objective values, but also their perceived importance [15].
- **Degree of grouping**: The degree of grouping of jobs of the same variant at subsequent positions in the sequence has a significant impact on several scheduling objectives such as flow time and makespan [12] depict the flow time as a function of the batch size with an optimum at medium degrees of grouping. Aloulou et al. [16] show that makespan is increased for high degrees of grouping. Felbecker [17] comes to similar results indicating that low and medium degrees of grouping are advantageous.

This work aims at analyzing the impacts of these four operational moderator variables on the relation between the schedule as independent variable and various scheduling objectives as dependent variables for mixed-model assembly lines. To analyze these relations, a structured simulation study is conducted based on data from a real mixed-model assembly line in the automotive industry. The results of these simulation studies are analyzed by testing several hypotheses on the impacts of the moderator variables.

Section 2 introduces the explored moderator variables as well as the scheduling objectives and describes the methodology used to analyze the impact of these operational moderating variables on the scheduling objectives. The results are presented in Sect. 3. Section 4 concludes this work and explores topics for future research.

2 Methodology to Analyze the Impact of the Moderator Variables

This work is based on the simulation study and the results presented in [18]. Within [18] the impact of five different sequence types on 13 scheduling objectives has been analyzed with a focus on the question which sequence type is preferred for which scheduling objective under which conditions. The conditions considered have been different number and heterogeneity of products, volume mixes, average workloads, and degrees of grouping within each analyzed sequence. This work does no longer set its focus on the preferred sequence types but on the impacts of these four moderator variables on the values of the scheduling objectives.

In the following subsections, the most important information on this work's study is given: in Sect. 2.1 on the scheduling objectives as dependent variables, in Sect. 2.2 on the sequences as independent variables and in Sect. 2.3 on the moderator variables. Section 2.4 finally describes how hypothesis on the gained simulation results are tested. For a more detailed description of the entire study settings, the case data, the simulation and the simulation results please refer to the previous work of [18].

2.1 Scheduling Objectives as Dependent Variables

The considered scheduling objectives can be grouped into productivity-, flow-, human-, supplier- and customer-related objectives and are summarized in Table 1. The formula to compute them are described in [18].

Table 1 Scheduling objectives

Objective category	Objective (Symbol)	Description
Productivity-related	Makespan (C_{max}) Throughput (TH)	= time span from the start time of a job set on the first station to its completion time on the last station = number of units completed within a scheduling period
Flow-related	Mean flow time (FTM) Flow time variance (FTV)	= average time-in-system of a job captures the time from the start on the first station to the completion on the last station = coefficient of variation of the single job flow times
Customer-related	Mean earliness (MEA)	= average time jobs are completed earlier than their due date
Human-related	Mean learning effect (MLE) Mean deterioration effect (MDE)	= mean gain per job in processing times compared to the normal processing times due to learning effects (e.g. routine gains) = mean increase per job in processing times compared to normal processing times due to deterioration effects (e.g. fatigue)
Supplier-related	Part usage rate variation (PUV)	= average hourly demand rate per part

2.2 Sequences as Independent Variables

In total, the impact of 570 different sequences on the scheduling objectives have been simulated. 95 sequences respectively can be grouped to each of the following sequence types:

- **No grouping**: Identical jobs are kept apart from each other with the maximum space possible between them within the sequence.
- **Minimum Part Set (MPS)**: The smallest set of jobs having the same volume mix proportion as the entire job set to be scheduled is repetitively produced. Within a MPS, the jobs are sorted according to the shortest processing time.
- **Groups of 4**: Identical jobs are grouped as packages of four within the sequence, with each job still being processed as a single job. These groups of four units are spaced maximally from each other in the sequence.
- **Groups of 12**: Analogously to groups of 4, sequences with groups of 12 are created.
- **Maximum grouping**: Identical jobs are successively positioned directly after another in the sequence. The job groups are sorted according to the shortest processing time.

2.3 Moderator Variables

All explored sequences of a sequence type vary systematically with respect to the moderator variables and each sequence contains information concerning the four moderator variables number and heterogeneity of products, volume mix, average workload and degree of grouping of identical jobs.

- Concerning **number and heterogeneity of products**, scenarios are introduced that differ in the product composition. The first two scenarios 3.1 and 3.3 encompass products 1 to 3 with similar workload and processing time to station distribution. The second two scenarios 5.1 and 5.5 include two additional low-workload products 4 and 5 with a differing processing time to station distribution but still the same number of stations used. The third two scenarios 7.1 and 7.6 include two additional high-workload products 6 and 7 that even require a higher number of stations used compared to the other five products. Thus, both number and heterogeneity of products increase over the scenarios. Every scenario consists of 95 explored sequences with 19 sequences analyzed for each sequence type.
- Within each of the six scenarios, the **volume mix percentages** are altered in a structured way. The volume percentage of a main product is gradually reduced from 95 to 5% in 5%-steps such that for every sequence type 19 sequences are explored. The remaining volume percentage is uniformly distributed between the other products considered in a scenario. The main products are product 1 in scenarios 3.1, 5.1 and 7.1, product 2 with a similar workload in scenario 3.2,

product 5 as product with the smallest workload in scenario 5.5 and product 6 as product with the highest workload in scenario 7.6.

- Within every scenario, **average workload** is systematically varied in dependence of the volume mix percentage. It stays quite constant in scenarios 3.1 and 3.2, while it increases in scenario 5.5 and decreases in scenarios 5.1, 7.1 and 7.6 with the volume mix percentage of the main product being reduced from 95 to 5%. The height of the average workload reaches its maximum in scenario 7.6 and its minimum in scenario 5.5.
- For each of the 19 volume mix points in every scenario, five different sequences are created that differ regarding the **degree of grouping**. The five sequences created belong to the sequence types presented in Sect. 2.3.

2.4 Hypothesis Testing

With each sequence containing information on the four moderator variables, the study setting allows to analyze the simulation results by means of hypothesis testing. Based on graphical analysis of the figures presented in [18], hypothesis are formulated that are then tested on the data set of the simulation results.

The hypotheses always follow the same formulation logic, for instance: *With a more heterogeneous volume mix, MFT decreases.* Several hypothesis have been tested on the simulation results for each moderator variable and the results are presented in Sect. 3.

3 Impact of Moderator Variables on Relation Between Schedule and Scheduling Objectives

3.1 Number and Heterogeneity of Products

For the number and heterogeneity of products, scenarios 3.1, 5.1 and 7.1 are compared as from 3.1 over 5.1 to 7.1 the number and heterogeneity of products increases under similar conditions. For every volume mix point of a sequence type, it is checked whether the objective value increases/decreases from scenario 3.1 over 5.1 to 7.1. The results are presented in Table 2.

What is already known from [18] is that the favorability among different sequencing types does not change but that the actual height of objective values is changed. This effect is particularly prevalent for MDE and MLE which show a strong dependency on the number and heterogeneity of products. For other objectives such as MEA and MFT the impact is not as straight-forward, but a tendency can still be observed.

Table 2 Impact of the number and heterogeneity of products

With a more heterogeneous product mix	In total (%)	3.1 to 5.1 (%)	3.1 to 7.1 (%)	5.1 to 7.1 (%)	No group (%)	MPS (%)	Groups of 4 (%)	Groups of 12 (%)	Max. group (%)
… *MDE* increases	**97.19**	94.74	98.95	97.89	98.25	98.25	96.49	92.98	100
… MLE decreases	**94.04**	94.74	100	87.37	91.23	92.98	89.47	96.49	100
… VFT increases	**79.30**	69.47	93.68	74.74	73.68	73.68	82.46	84.21	82.46
… PUV increases	**70.53**	95.79	93.68	22.11	82.46	66.67	75.44	63.16	64.91
…C_{max} increases	**70.18**	61.05	81.05	68.42	78.95	80.70	77.19	80.70	33.33
… TH decreases	**64.56**	64.21	70.53	58.95	84.21	80.70	78.95	64.91	5.26
… MFT increases	**64.56**	45.26	74.74	73.68	75.44	68.42	71.93	75.44	31.58
… MEA increases	**62.46**	68.42	68.42	50.53	56.14	52.63	59.65	64.91	78.95

3.2 Volume Mix

For the volume mix, the objective values of every sequence type within a scenario are compared for every pair of adjacent volume mix points. The results are presented in Table 3.

In general, homogeneous volume mixes with a dominating product in the mix often lead to other objective values compared to heterogeneous volume mixes that allow using alternation effects between various products in a sequence. A strong dependency is observed especially for MFT and MLE both decreasing with a more heterogeneous volume mix. For other objectives like PUV and VFT, the relationship is less strong, in particular for sequence types with a low degree of grouping.

3.3 Degree of Grouping

For the degree of grouping, the objective values of sequence types with a different degree of grouping are compared within a scenario for the same volume mix point. Please note that the results for the MPS are not considered as it cannot be clearly stated whether MPS leads to a higher or lower degree of grouping especially compared to other sequence types with a low degree of grouping. The results are presented in Table 4.

With a sequence with a higher degree of grouping, PUV and MLE almost always increase and VFT decreases. For the other objectives, at least a tendency can be observed except for makespan for which not even a tendency is observable.

3.4 Workload

For average workload, for every sequence type the objective values for adjacent volume mix points within a scenario are compared. The results can be found in Table 5. In comparison to the other moderator variables, lower dependency values are observed for average workload. Average workload plays a less important but still recognizable role as moderator variable for most objectives except for throughput and mean flow time for which workload is a major explanatory factor.

4 Conclusion and Future Research

The impacts of moderator variables on the relationships between schedules and scheduling objectives are not straight-forward. The presented results show that at least

Table 3 Impact of the volume mix

With a more heterogeneous volume mix	In total (%)	3.1 (%)	3.2 (%)	5.1 (%)	5.5 (%)	7.1 (%)	7.6 (%)	No group (%)	MPS (%)	Groups of 4 (%)	Groups of 12 (%)	Max. group (%)
… MFT decreases	**78.52**	91.11	58.89	97.78	53.33	94.44	75.56	74.07	75.93	79.63	86.11	76.85
… MLE decreases	**77.78**	62.22	67.78	87.78	78.89	87.78	82.22	80.56	82.41	81.48	78.70	65.74
… MDE increases	**72.04**	54.44	52.22	81.11	75.56	86.67	82.22	74.07	75.00	70.37	78.70	62.04
… C_{max} decreases	**70.74**	77.78	63.33	75.56	41.11	72.22	94.44	69.44	66.67	70.37	72.22	75.00
… TH increases	**70.56**	68.89	87.78	72.22	27.78	66.67	100	67.59	69.44	69.44	69.44	76.85
… MEA increases	**68.89**	54.44	56.67	74.44	60.00	78.89	88.89	69.44	67.59	68.52	67.59	71.30
… PUV increases	**62.04**	77.78	31.11	78.89	80.00	82.22	22.22	59.26	61.11	62.96	59.26	67.59
… VFT increases	**61.30**	72.22	42.22	72.22	38.89	73.33	68.89	51.85	53.70	62.96	60.19	77.78

Table 4 Impact of the degree of grouping

With a higher degree of grouping	In total (%)	No to 4 (%)	No to 12 (%)	No to max (%)	4 to 12 (%)	4 to max (%)	12 to max (%)	3.1 (%)	3.2 (%)	5.1 (%)	5.5 (%)	7.1 (%)	7.6 (%)
… PUV increases	**99.71**	100	100	100	99.12	100	99.12	100	99.12	99.12	100	100	100
… MLE increases	**98.98**	97.37	99.12	100	97.37	100	100	100	99.12	99.12	100	98.25	97.37
… VFT increases	**94.59**	91.23	94.74	98.25	91.23	99.12	92.98	97.37	92.11	99.12	91.23	96.49	91.23
… MDE decreases	**86.84**	69.30	84.21	100	82.46	97.37	87.72	73.68	71.93	91.23	94.74	93.86	95.61
… MEA increases	**74.27**	88.60	94.74	70.18	92.11	56.14	43.86	52.63	50.88	72.81	93.86	76.32	99.12
… TH increases	**66.81**	45.61	89.47	70.18	86.84	64.04	44.74	41.23	36.84	69.30	81.58	83.33	88.60
… MFT increases	**62.13**	29.82	56.14	70.18	62.28	73.68	80.70	76.32	86.84	48.25	77.19	64.04	20.18
… C_{max} increases	**49.56**	35.96	21.93	66.67	22.81	70.18	79.82	64.04	73.68	47.37	45.61	51.75	14.91

Table 5 Impact of workload

With higher average workload	In total (%)	3.1 (%)	3.2 (%)	5.1 (%)	5.5 (%)	7.1 (%)	7.6 (%)	No group (%)	MPS (%)	Groups of 4 (%)	Groups of 12 (%)	Max. group (%)
… TH decreases	**74.81**	68.89	46.67	72.22	94.44	66.67	100	71.30	71.30	72.22	73.15	86.11
… MFT increases	**74.44**	91.11	41.11	97.78	46.67	94.44	75.56	74.07	70.37	74.07	71.30	82.41
…C_{max} increases	**69.26**	77.78	36.67	75.56	58.59	72.22	94.44	65.74	62.96	64.81	66.67	86.11
… VFT decreases	**67.59**	72.22	57.78	72.22	61.11	73.33	68.89	66.67	66.67	64.81	67.59	72.22
… MEA decreases	**63.33**	54.44	43.33	74.44	40.00	78.89	88.89	62.04	62.04	61.11	63.89	67.59
… MDE decreases	**62.78**	54.44	47.78	81.11	24.44	86.67	82.22	64.81	65.74	62.96	65.74	54.63
… MLE increases	**62.22**	62.22	32.22	87.78	21.11	87.78	82.22	60.19	62.04	62.96	62.04	63.89
… PUV increases	**58.33**	77.78	68.89	78.89	20.00	82.22	22.22	57.41	55.56	51.85	59.26	67.59

some strong relationships exist. A more heterogeneous volume mix helps to decrease the mean flow time, yet it comes at costs of smaller mean learning effects. Higher degrees of grouping lead to higher part usage rate variations, higher mean learning effects per job and a higher flow time variance. With increasing average workload, throughput decreases and mean flow time increases. The number and heterogeneity of products particularly impacts mean learning and deterioration effects.

This contribution provides merely a first step in the exploration of the impact of operational moderator variables on the relationship between schedules and scheduling objectives. The relationships have to be analyzed for data sets of other mixed-model assembly lines in order to assess whether the observed relationships also hold under different conditions. Moreover, this study indicates that there is seldom one dominating moderator variable but that all four investigated moderator variables have at least some explanatory powers. Hence, multiple regression models need to be used to explore their joint explanatory power.

References

1. Baron, R.M., Kenny, D.A.: The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol. **51**(6), 1173–1882 (1986)
2. Allen, M.: The efficient utilization of labor under conditions of fluctuating demand. In Muth, J.F., Thompson, G.L.: Industrial Scheduling. Prentice-Hall, Englewood Cliffs, 252–276 (1963)
3. Boysen, N., Fliedner, M., Scholl, A.: Produktionsplanung bei Variantenfließfertigung: Planungshierarchie und Elemente einer Hierarchischen Planung. J. Bus. Econ. **77**(7–8), 759–793 (2007)
4. Boysen, N.: Variantenfließfertigung. Gabler, Wiesbaden (2005)
5. Freiheit, T., Li, W.: The effect of work content imbalance and its interaction with scheduling method on sequential flow line performance. Int. J. Prod. Res. **55**(10), 2791–2805 (2017)
6. Scholl, A.: Balancing and sequencing of assembly lines. Ph.D. Thesis, Technische Universität Darmstadt (1999)
7. El-Rayah, T.E.: The efficiency of balanced and unbalanced production lines. Int. J. Prod. Res. **17**(1), 61–75 (1979)
8. Boysen, N., Fliedner, M., Scholl, A.: Production planning of mixed-model assembly lines: overview and extensions. Prod. Planning Control **20**(5), 455–471 (2009)
9. Hottenstein, M.P., Bowman, S.A.: Cross-training and worker flexibility: A review of DRC system research. J. High Technol. Managem. Res. **9**(2), 157–174 (1998)
10. Alavian, P., Denno, P., Meerkov, S.M.: Multi-job production systems: definition, problems, and product-mix performance portrait of serial lines. Int. J. Prod. Res. **55**(24), 7276–7301 (2017)
11. Anderson, S.W.: Direct and indirect effects of product mix characteristics on capacity management decisions and operating performance. Int. J. Flex. Manuf. Syst. **13**(3), 241–265 (2001)
12. Hopp, W.J., Spearman, M.L.: Factory Physics. Waveland Press, Long Grove, IL (2011)
13. Framinan, J., Leisten, R., Ruiz, R.: Manufacturing Scheduling Systems: An Integrated View of Models, Methods, and Tools. Springer-Verlag, London (2014)
14. Eilon, S., Cotterjll, D.J.: A modified SI rule in job shop scheduling. Int. J. Prod. Res. **7**(2), 135–145 (1968)
15. Pinedo, M.L.: Scheduling: Theory, Algorithms, and Systems, 5th edn. Springer, New York, NY (2016)

16. Aloulou, M.A., Bouzaiene, A., Dridi, N., Vanderpooten, D.: A bicriteria two-machine flow-shop serial-batching scheduling problem with bounded batch size. J. Sched. **17**(1), 17–29 (2014)
17. Felbecker, O.: Ein Beitrag zur Reihenfolgeplanung bei Mehrprodukt-Linienfertigung. Ph.D. Thesis, RWTH Aachen (1980)
18. Ostermeier, F.F.: On the trade-offs between scheduling objectives for unpaced mixed-model assembly lines. Int. J. Prod. Res. 1–28 (2020)

Proposal of the Algorithm for Solving the Component Assignment Problem in a Linear Consecutive System with Three Failure Modes

Taishin Nakamura

Abstract Recently, a dual linear consecutive system with three failure modes was proposed. This system contains two parallel subsystems consisting of n components arranged in a line. The system fails if (1) subsystem 1 has at least k_1 consecutive failed components, (2) subsystem 2 has at least k_2 consecutive failed components, or (3) the system has m consecutive pairs of failed components. This system has diverse applications, such as road lights on highways. To design a reliable system, this study addresses the problem of finding the component arrangement that maximizes the system reliability, assuming that the components are functionally exchangeable. A simple enumeration method can theoretically find the optimal arrangement, but it is computationally expensive and impracticable for larger numbers of components. Furthermore, no algorithm for efficiently finding the optimal arrangement of a dual linear consecutive system with three failure modes has been established. Thus, this study develops a branch-and-bound-based algorithm to solve the component assignment problem for a dual linear consecutive system with three failure modes. For an efficient reliability computation in the proposed algorithm, we combine a recursive equation with the branch-and-bound method. This results in the elimination of redundant computations. Furthermore, we derive a condition for pruning based on system reliability, which can reduce the number of enumerated arrangements. Through numerical experiments, we confirm the usefulness of eliminating redundant computations and pruning based on system reliability by comparing the proposed algorithm with the enumeration method. The proposed algorithm can improve the reliability of real-world systems expressed as dual linear consecutive systems with three failure modes.

Keywords System reliability · Component assignment problem · Branch-and-bound method

T. Nakamura (✉)
Tokai University, Kanagawa, Japan
e-mail: nakamura@tsc.u-tokai.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_41

1 Introduction

With technological advancements in various fields, the number of components in related systems is increasing, and their structure is becoming more sophisticated. As a result, practical systems are becoming increasingly complex, and it is increasingly difficult to ensure the reliability of such systems. Reliability is a major attribute used to evaluate the performance of practical systems. It is generally defined as the probability that a system will perform satisfactorily for at least a given period of time under a specific condition of use.

In recent decades, researchers have aimed to maximize the reliability of various systems with different structures. Various reliability models have been proposed for practical systems. Such system models can be classified into two categories: those with a single failure criterion and those with multiple failure criteria. A consecutive-k-out-of-n: F system has a single system failure criterion. This system consists of n components arranged in a line and fails when k consecutive components fail. Salvia and Lasher [1] extended this system to two dimensions, i.e., a connected-(r, s)-out-of-(m, n): F lattice system. This system consists of mn components arranged in an $m \times n$ lattice; it fails if there is at least one $r \times s$ matrix in which all components fail.

Systems with multiple failure criteria have also been proposed. Tung [2] presented an (n, f, k) system. This system consists of n components arranged in a line and fails if there are at least f failed components or at least k consecutive failed components. Cui et al. [3] proposed an $\langle n, f, k \rangle$ system. This system consists of n components arranged in a line and fails if there are at least f failed components and at least k consecutive failed components. In addition, Boehme et al. [4] defined a two-dimensional system with multiple failure criteria. A connected-(1, 2)-or-(2, 1)-out-of-(m, n): F lattice system, which comprises components arranged in an $m \times n$ lattice, fails if two or more consecutive failed components exist in any one row or column.

Recently, Peng et al. [5] proposed a dual linear consecutive system with three failure modes. This system consists of two consecutive-k-out-of-n: F subsystems arranged in parallel. As shown in Fig. 1, the system fails if (1) subsystem 1 contains at least k_1 consecutive failed components, (2) subsystem 2 contains at least k_2 consecutive failed components, or (3) the system has m consecutive pairs of failed components, where $m \leq k_1$ and k_2. For simplicity, the dual linear consecutive system with three failure modes is referred to as the DLC $(k_1, k_2, m)/n$ system in this paper, where k_1, k_2, and m are failure criteria parameters. DLC $(k_1, k_2, m)/n$ systems have diverse applications in many fields. For example, consider a highway that needs road lights on both sides during nighttime. Road lights are installed at regular intervals, e.g., 20 m. When a certain number of consecutive lights on the same side are broken, there will be insufficient light in that area. Additionally, consecutive pairs of failed lights on both sides can lead to a dark area. Such situations may affect the traffic. Thus, such road lights can be modeled as a DLC $(k_1, k_2, m)/n$ system.

The main purpose of reliability optimization is to design a reliable system to prevent accidents and reduce the causes of failure. Reliability optimization has recently received considerable attention in academic and industrial research. This

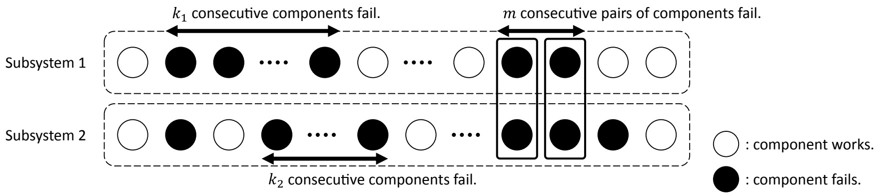


Fig. 1 Three failure modes of a DLC $(k_1, k_2, m)/n$ system

study addresses the component assignment problem (CAP) as a reliability optimization problem. Assuming that components are functionally exchangeable, different component arrangements will result in different system reliabilities. Hence, the system reliability can be improved by optimally assigning components to positions in the system. Accordingly, this problem involves finding the component arrangement that maximizes system reliability. Determining the optimal arrangement will contribute to designing reliable systems.

A simple enumeration method can theoretically determine the optimal arrangement. Specifically, this method enumerates all of the arrangements (permutations), evaluates the reliability of the system for each arrangement, and ultimately determines the optimal component arrangement that achieves the highest reliability. However, this method is computationally expensive and thus impracticable when the number of components is not small. Furthermore, no study has established an algorithm for efficiently determining the optimal arrangement of DLC $(k_1, k_2, m)/n$ systems.

The aim of this study is to develop an exact algorithm for efficiently solving the CAP for DLC $(k_1, k_2, m)/n$ systems. To decrease the computational burden, we provide a reliability computation method that can be regarded as an alternate expression of the existing method [5]. The combination of the proposed method and algorithm can achieve efficient computation by eliminating redundant calculations. Furthermore, using the system reliability, we reduce the number of enumerated arrangements; specifically, arrangements that cannot be optimal are removed. Hence, the proposed algorithm is based on the branch-and-bound (B&B) approach. Finally, the efficiency of the proposed algorithm is investigated through computational experiments.

2 CAP in the Dual Linear Consecutive System with Three Failure Modes

2.1 Assumptions, Notation, and Formulation

First, we assume that (1) each component and the system can have only two states: working or failed; (2) all components are mutually statistically independent; (3)

the component reliabilities are given; and (4) the components are functionally interchangeable.

The symbols and notations used in this paper are as follows. For $i = 1, 2$ and $j = 1, 2, \ldots, n$, let (i, j) be the j-th position of subsystem i. The symbol $\pi(i, j)$ represents an index of the component assigned to position (i, j) for $i = 1, 2$ and $j = 1, 2, \ldots, n$, where $\pi(i, j) \in \{1, 2, \ldots, 2n\}$. In addition, an arrangement of mn components in which component $\pi(i, j)$ is assigned to position (i, j) is given by

$$\Pi = \begin{pmatrix} \pi(1,1) & \pi(1,2) & \cdots & \pi(1,n) \\ \pi(2,1) & \pi(2,2) & \cdots & \pi(2,n) \end{pmatrix}_{2\times n}. \tag{1}$$

For $\tau \in \{1, 2, \ldots, 2n\}$, let τ be an index of the component; the reliability of component τ is p_τ, where $p_1 < p_2 < \ldots < p_{2n}$ without loss of generality. Hence, p_τ indicates the reliability of the τ-th least reliable component. The vector $\boldsymbol{p}$ represents a $2n$-vector of the component reliabilities $(p_1, p_2, \ldots, p_{2n})$. Here, if $R(n; (k_1, k_2, m), \boldsymbol{p}; \Pi)$ is the reliability of the DLC $(k_1, k_2, m)/n$ system under arrangement Π with component reliabilities $\boldsymbol{p}$, then the CAP is to determine an arrangement that maximizes the reliability of the DLC $(k_1, k_2, m)/n$ system (Π^*), i.e.,

$$\Pi^* = \text{argmax}_{\Pi\in\Omega} R(n; (k_1, k_2, m), \boldsymbol{p}; \Pi), \tag{2}$$

where Ω is a set of all arrangements.

2.2 Related Studies

This subsection presents previous studies on the DLC $(k_1, k_2, m)/n$ system. Peng et al. [5] introduced a DLC $(k_1, k_2, m)/n$ system and derived a recursive equation to compute the system reliability.

Generally, newly introduced reliability system models often coincide with conventional models under certain restrictions of the design parameters [6]. For example, a DLC $(k_1, k_2, m)/n$ system with $(k_1, k_2, m) = (2, 2, 1)$ becomes a connected-(1, 2)-or-(2, 1)-out-of-$(2, n)$: F lattice system. Recently, Sui et al. [7] investigated a property that can be used as a necessary condition that optimal arrangements must satisfy. Moreover, they developed an algorithm to determine the optimal arrangement of a connected-(1, 2)-or-(2, 1)-out-of-(m, n): F lattice system. Furthermore, Hwang and Shi [8] investigated the properties of the optimal arrangement of a connected-(2, 1)-out-of-$(2, n)$: F lattice system, and Omura et al. [9] developed an algorithm to determine the optimal arrangement of a connected-$(m, 2)$-out-of-(m, n): F lattice system. Subsequently, Nakamura and Yamamoto [10] developed an algorithm to determine the optimal arrangement of general connected-(r, s)-out-of-(m, n): F lattice systems. Furthermore, Several studies have aimed to obtain the optimal arrangement of components for the above systems. However, an algorithm for determining the optimal arrangement of DLC $(k_1, k_2, m)/n$ systems has not yet been reported. Thus, this study develops an algorithm for efficiently obtaining the optimal arrangement of a DLC $(k_1, k_2, m)/n$ system.

3 Recursive Equation to Compute the System Reliability

This section presents a recursive equation for computing the reliability of the DLC $(k_1, k_2, m)/n$ system. Here, system reliability is the objective function of the CAP. Peng et al. [5] derived a recursive equation to compute the reliability of DLC $(k_1, k_2, m)/n$ systems. However, this study proposes a new recursive equation as an alternative expression of the existing equation [5].

The concept of the system reliability computation is to decompose the DLC $(k_1, k_2, m)/n$ system into several smaller systems in which some components have certain conditions. Consequently, the reliability of the DLC $(k_1, k_2, m)/n$ system can be expressed recursively as the summation of the reliabilities of the smaller systems. Here, for $j = 1, 2, \ldots, n$; $h_1 = 0, 1, \ldots, k_1$; and $h_2 = 0, 1, \ldots, k_2$, let $R(j; (h_1, h_2))$ denote the probability that the following three conditions are satisfied:

(1) The DLC $(k_1, k_2, m)/j$ system does not have any failure modes;
(2) In subsystem 1, components $\pi(1, j - h_1 + 1), \pi(1, j - h_1 + 2), \ldots, \pi(1, j)$ fail, and $\pi(1, j - h_1)$ works;
(3) In subsystem 2, components $\pi(2, j - h_2 + 1), \pi(2, j - h_2 + 2), \ldots, \pi(2, j)$ fail, and $\pi(2, j - h_2)$ works.

Figure 2 illustrates the smaller system in which some components have conditions h_1 and h_2.

Using the definition of the smaller system, we compute the reliability of the DLC $(k_1, k_2, m)/n$ system using the following recursive equation.

Theorem 1 (a) For $j = 1, 2, \ldots, n$; $h_1 = 0, 1, \ldots, k_1$; and $h_2 = 0, 1, \ldots, k_2$,

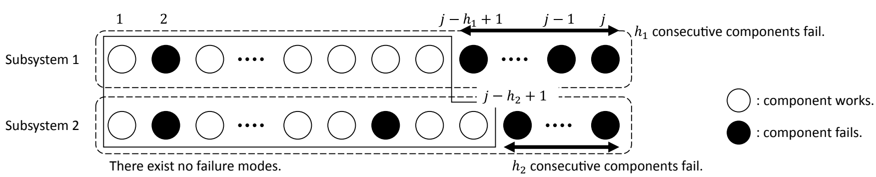


Fig. 2 Smaller system in which some components have conditions h_1 and h_2

$$R(j;(h_1,h_2)) = \begin{cases} p_{\pi(1,j)}p_{\pi(2,j)} \sum_{(g_1,g_2)\in A} R(j-1;(g_1,g_2)) & h_1 = 0 \text{ and } h_2 = 0, \\ p_{\pi(1,j)}\left(1-p_{\pi(2,j)}\right) \sum_{0\le g_1\le k_1-1} R(j-1;(g_1,h_2-1)) & h_1 = 0 \text{ and } 0 < h_2 < k_2, \\ \left(1-p_{\pi(1,j)}\right)p_{\pi(2,j)} \sum_{0\le g_2\le k_2-1} R(j-1;(h_1-1,g_2)) & 0 < h_1 < k_1 \text{ and } h_2 = 0, \\ \left(1-p_{\pi(1,j)}\right)\left(1-p_{\pi(2,j)}\right)R(j-1;(h_1-1,h_2-1)) & 0 < h_1 < k_1 \text{ and } 0 < h_2 < k_2, \\ 0 & otherwise, \end{cases} \tag{3}$$

where

$$A = \{(h_1,h_2)|\min\{h_1,h_2\} < m \text{ for } h_1 = 0,1,\ldots,k_1 - 1 \text{ and } h_2 = 0,1,\ldots,k_2-1\}. \tag{4}$$

As the boundary condition, for,

$$R(0;(h_1,h_2)) = \begin{cases} 1 & (h_1,h_2) = (0,0), \\ 0 & otherwise. \end{cases} \tag{5}$$

(b)

$$R(n;(k_1,k_2,m),\boldsymbol{p};\Pi) = \sum_{(h_1,h_2)\in A} R(n;(h_1,h_2)). \tag{6}$$

The main difference between the recursive equations derived in here and the existing recursive equations [5] is the way the DLC $(k_1, k_2, m)/n$ system is decoosed. In the derived recursive equation, the system is decomposed into smaller systems in which some components have specific conditions. However, in the existing recursive equations [5], the system is decomposed into smaller subsystems with different sizes. This difference results in a relatively simple summation of the derived recursive equations.

4 Proposed B&B-Based Algorithm

Herein, we present a B&B-based algorithm for determining the optimal arrangement of a DLC $(k_1, k_2, m)/n$ system. The framework of the proposed algorithm is similar to that in Nakamura and Yamamoto [9] for finding the optimal arrangement of a connected-(r, s)-out-of-(m, n): F lattice system.

The key concept of a B&B algorithm is detecting and discarding candidates that cannot be part of the optimal arrangement. Partial arrangements that do not lead to an optimal arrangement can be eliminated from further searches. This elimination corresponds to pruning the branches of the search tree. Incorporating the conditions for pruning accelerates the B&B algorithm by reducing the number of candidates for the optimal arrangement. In the proposed algorithm, when enumerating the component arrangements, we assign the components in the order of positions (1, 1), (2, 1), (2, 1), (2, 2), . . . , (n, 1), (n, 2) in the depth-first search tree. To find the optimal arrangement efficiently, pruning should be performed as many times as possible. This algorithm includes the following two conditions:

(1) Pruning to eliminate symmetrical arrangements; and
(2) Pruning based on system reliability.

The reliabilities of systems with symmetrical arrangements are the same. For example, the arrangement obtained by rotating the left and right ends of the original system will have the same system reliability. Hence, only one symmetrical arrangement is retained, which can gradually reduce the number of candidates and improve the efficiency of obtaining the optimal arrangement. We apply the following conditions to eliminate all symmetrical arrangements except one.

Condition 1

(a) For $k_1 = k_2$:
When a component is assigned to position (2, 1), pruning is performed if $\pi(1, 1) < \pi(2, 1)$ is not satisfied.
(b-1) When the number of components in each subsystem (n) is even:
When a component is assigned to position $\left(1, \frac{n}{2} + 1\right)$, pruning is performed if $\pi\left(1, \frac{n}{2}\right) < \pi\left(1, \frac{n}{2} + 1\right)$ is not satisfied.
(b-2) When the number of components in each subsystem (n) is odd:
When a component is assigned to position $\left(1, \frac{n+1}{2} + 1\right)$, pruning is performed if $\pi\left(1, \frac{n+1}{2} - 1\right) < \pi\left(1, \frac{n+1}{2} + 1\right)$ is not satisfied.

Combining the B&B method with the recursive equation in Theorem 1 can enhance the performance of the proposed algorithm. First, the recursive equation can reduce the time complexity of the reliability computations. When searching for the optimal arrangement, we compute the reliabilities of DLC $(k_1, k_2, m)/n$ systems with common partial arrangements several times. Memorization allows the storage of the previously computed system reliabilities to eliminate redundant computations, thus drastically reducing the time complexity.

The algorithm then decides whether to prune the current branch or assign a further component based on an upper bound of the reliability for a partial system. If the upper bound does not exceed the incumbent value (the maximum system reliability obtained so far), then further assignment of components cannot lead to a solution better than the incumbent value.

Condition 2
When a component is assigned to position $(2, j)$ $(j = 1, 2, \ldots, n)$, pruning is performed if the following is not satisfied:

$$\mathbf{Rmax} \leq \sum_{(h_1,h_2)\in A} R(j; (h_1, h_2)), \tag{7}$$

where **Rmax** is the incumbent value (the maximum system reliability obtained thus far), and set A is given by Eq. (4).

It is evident from the recursive equation derived in Sect. 3 that the system reliability is a monotonically non-increasing function with respect to the number of assigned components.

5 Efficiency Investigation

This section presents the results of numerical experiments to evaluate the effectiveness of the proposed algorithms. The performance criterion was the computation time required to obtain a solution. The experiments were performed in a program written in the C++ programming language and compiled with Visual Studio 2019. All experiments were performed on a computer with an Intel(R) Core(TM) i7-9700 (3.00 GHz) CPU with 8.00 GB RAM and Microsoft Windows 10. The component reliabilities $p_\tau (\tau = 1, 2, \ldots, 2n)$ were randomly generated from a uniform distribution on [0.00, 1.00]. The numerical experiment results are the average computation times for 10 instances with different random number seeds.

Herein, we compare the proposed algorithm with the enumeration method. The proposed algorithm computes the system reliability for each j, whereas the enumeration method computes it after all of the components are assigned. As a result, the enumeration method does not include Condition 2. Table 1 lists a comparison of the computation times required to find the optimal arrangements of DLC $(3, 3, 2)/n$ systems; N/A indicates a failure to obtain an optimal arrangement in 48 h. Because the proposed algorithm becomes more advantageous as the number of components in each subsystem increases, we conclude that eliminating redundant computations and pruning based on the system reliability are useful for efficiently obtaining the optimal arrangement.

Table 1 Computation times required to find the optimal arrangements of DLC $(3, 3, 2)/n$ systems [s]

n (# of components in each subsystem)	5	6	7	8
(A) proposed algorithm	0.02	3.13	160.41	21,096.21
(B) enumeration method	3.45	542.43	114,762.18	N/A
Ratio: (B)/(A)	154.67	173.53	715.43	– -

6 Conclusion

In this study, we developed a B&B-based algorithm that can efficiently solve the CAP for DLC $(k_1, k_2, m)/n$ systems. For an efficient reliability computation in the proposed algorithm, we provided a recursive equation combined with the B&B method that can eliminate redundant computations. Furthermore, we derived a pruning condition based on the system reliability. Through numerical experiments, we confirmed the usefulness of eliminating redundant computations and pruning based on the system reliability by comparing the proposed algorithm with the enumeration method. The proposed algorithm can help improve the reliability of real-world systems expressed as DLC $(k_1, k_2, m)/n$ systems.

Generally, the efficiency of solving large-scale combinatorial optimization problems can be enhanced by utilizing the properties of each problem. Further studies could derive pruning rules for a B&B method based on the properties of the CAP for DLC $(k_1, k_2, m)/n$ systems, which would improve the efficiency of the proposed algorithm.

Acknowledgements This work was supported by the Nakajima Foundation (http://www.nakajimafound.or.jp/) and Grant-in-Aid for Early-Career Scientists, Grant Numbers JP21K14370.

References

1. Salvia, A.A., Lasher, W.C.: Two-dimensional consecutive-*k*-out-of-*n*: f models. IEEE Trans. Reliab. **39**(3), 382–385 (1990). https://doi.org/10.1109/24.103023
2. Tung, S.S.: Combinatorial analysis in determining reliability. In: Proceedings of the Annual Reliability and Maintainability Symposium, pp. 262–266 (1982)
3. Cui, L., Kuo, W., Li, J., Xie, M.: On the dual reliability systems of (n, f, k) and $< n, f, k$. Statist. Probab. Lett. **76**(11), 1081–1088 (2006). https://doi.org/10.1016/j.spl.2005.12.004
4. Boehme, T.K., Kossow, A., Preuss, W.: A generalization of consecutive-*k*-out-of-*n*: f systems. IEEE Trans. Reliab. **41**(3), 451–457 (1992). https://doi.org/10.1109/24.159819
5. Peng, R., Wu, D., Gao, K.: Reliability of a dual linear consecutive system with three failure modes. In: International Conference Celebrating Professor Jinhua Cao's 80th Birthday, pp. 259–269 (2019)
6. Triantafyllou, I.S.: Consecutive-type reliability systems: an overview and some applications. J. Qual. Reliab. Eng. **2015**, 20 (2015). https://doi.org/10.1155/2015/212303
7. Sui, T., Nakamura, T., Yamamoto. H.: Investigating properties of the optimal arrangement in connected-(1,2)-or-(2,1)-out-of-(*m*, *n*): f systems. Tech. Report, IEICE (2021)
8. Hwang, F.K., Dinghua, S.: Redundant consecutive- *k* systems. Oper. Res. Lett. **6**(6), 293–296 (1987). https://doi.org/10.1016/0167-6377(87)90046-0
9. Omura, T., Akiba, T., Xiao, X., Yamamoto, H.: Algorithm for obtaining optimal arrangement of a connected-(*r*, *s*)-out-of-(*m*, *n*): f system—the case of $m = r$ and $s = 2$—. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **98**(10), 2018–2024 (2015). https://doi.org/10.1587/transfun.E98.A.2018
10. Nakamura, T., Yamamoto, H.: Algorithm for solving optimal arrangement problem in connected-(*r*, *s*)-out-of-(*m*, *n*): f lattice system. IEEE Trans. Reliab. **69**(2), 497–509 (2019). https://doi.org/10.1109/TR.2019.2925142

Metaheuristics Based Profit-Oriented Optimization Model for the Hazardous Waste Location Routing Problem

Amani Junaidi and Takashi Irohara

Abstract Hazardous Waste Management is defined as the safe and efficient handling of hazardous waste to decrease its toxicity to humans and the environment by means of proper transportation, processing, and disposal. Hazardous waste is generated by different sectors, including the medical sector, industrial sector, domestic sector, among others. Industrial hazardous waste consists of chemicals and compounds of a complex structure, which poses a great risk for public health and the environment alike. This potential risk heavily influences decision-makers when choosing a suitable location for establishing a waste processing facility. Therefore, the process of designing a hazardous waste transportation network comes with many challenges. This paper presents a profit-oriented mixed-integer linear programming model for the hazardous waste location-routing problem, with the main objective of maximizing the overall profit in the network and conditions focusing on minimizing the associated risk in terms of population exposure. The transportation network includes waste generators which are, in this case, factories, along with three different types of hazardous waste processing centers: treatment, recycling, and disposal centers. The formulated problem is coded in Python and optimally solved by Gurobi Optimizer. Furthermore, to deal with the NP-hard nature of the problem for large numerical instances, a metaheuristics algorithm based on non-dominated sorting genetic algorithm II (NSGA-II) is applied, and the model is solved with both NSGA-II and Gurobi to investigate the improvements done by utilizing the genetic algorithm. A case study is conducted for the textile industry in Jordan, as to put the proposed model into practice.

Keywords Hazardous waste management · Location-routing problem · Metaheuristics

---

A. Junaidi (✉) · T. Irohara
Sophia University, Tokyo, Japan
e-mail: a-junaidi@eagle.sophia.ac.jp

T. Irohara
e-mail: irohara@sophia.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_42

1 Introduction

Hazardous Waste Management (HWM) has been a topic of concern in recent years. Hazardous materials (hazmat) generated by the industrial sector remain as one of the biggest threats to humans and the environment, as most manufacturers are located in developing countries with loose regulations on waste management and disposal.

Research on hazardous waste transportation has been mostly focused on the design of networks on a strategic level through the Facility Location Problem (FLP). Nonetheless, there have been many recent research efforts addressing HWM on an operational level through the Vehicle Routing Problem (VRP). The Location Routing problem (LRP) tackles the design of a transportation network on both strategic and operational levels. In this paper, we present a profit-oriented mixed-integer linear programming model (MILP) for the hazardous waste location-routing problem, with conditions focusing on the associated risk. The transportation network consists of factories along with three different types of waste processing centers: treatment, recycling, and disposal centers.

In general, there are two main sources of profit in hazmat transportation networks: income from the polluter pays principle, and income from sales of electricity generated by energy recovery [1]. However, this paper focuses on the profit generated from the repurposing of recycled hazardous waste, by reusing it in manufacturing processes or selling it commercially. In the case of developing countries, this would be an appealing approach to encourage companies to invest in waste recycling, by offering a new lens on how profit can be generated.

The formulated problem is coded in Python and optimally solved by Gurobi Optimizer. A non-dominated sorting genetic algorithm II (NSGA-II) is proposed to tackle the NP-hard nature of the problem and obtain the non-dominated Pareto solution in a reasonable time. Lastly, a case study is conducted for the textile industry in the North and Central regions of Jordan, as to put the proposed model into practice.

2 Literature Review

Interest in hazmat logistics has been rapidly increasing over the past two decades. One of the earliest known literature efforts in hazmat transportation management is Glickman [2] where they address the problem of railroad shipments of hazardous materials with the purpose of avoiding populated areas.

Holeczek [3] published a review on hazmat truck transportation problems where the following classification was proposed according to the contribution of the papers: (a) Risk assessment (b) Routing (c) Routing and location (d) Network design (e) Toll setting. Risk assessment and routing problems made up most of the research content, while LRP made up only 10% of the 290 papers included in the survey.

There have been many recent efforts to expand the research done on hazmat LRP. Zografos and Samara [4] remains one of the earliest studies where a multi-objective model was proposed aiming to minimize transportation risk, travel time, and disposal risk. Wang et al. [5] constructed a nonlinear integer open location-routing model for relief distribution considering travel time, total cost, and reliability. Beneventti et al. [6] considered different types of hazmat in a multi-objective model that maximizes the minimum weighted distance between hazardous facilities and the exposed population.

Heuristics are also heavily employed in hazmat LRP. Martínez-Salazar et al. [7] focused on solving the bi-objective LRP by metaheuristic algorithms by reduction of distribution cost and balance of workloads. Pichka et al. [8] addressed the two-echelon open location routing problem with hybrid heuristic, while Nedjati et al. [9] investigated a bi-objective integer linear programming model that minimizes the waiting time and lost demands.

When it comes to research focusing on potential profit, Boyer et al. [10] proposed the idea of minimizing the cost of the hazmat transportation network by considering the income from selling recycled waste, while Aydemir-Karadag [1] integrated the polluter pays principle and income from sales of electricity generated by energy recovery as means to maximize the profit.

Our proposed model considers both the profit generated by selling recycled waste, as well as the annual profit of factories as an indication of how the overall profit of the network is affected. To the best of our knowledge, our approach of tackling the overall profit in the hazmat transportation network is yet to be studied.

3 Problem Description and Formulation

3.1 Problem Framework

The proposed hazardous waste transportation network is shown in Fig. 1, where the generation points of hazardous waste are factories. In the first stage, the generated waste is classified into three main categories: (1) Recyclable (2) Treatable (3) Non-recyclable and non-treatable. In accordance with that, the waste gets transported to either a recycling center, treatment center, or a disposal center and in different directions.

In the case of waste transported to recycling centers, there are three possible outcomes: (1) Successfully recycled waste that can be repurposed or sold (2) Recycled waste that cannot be repurposed and goes to industrial landfills as non-hazardous residue (3) Waste that is not successfully recycled and gets transported to hazardous waste disposal centers.

In the case of treatment centers, the waste that is successfully treated can either be transported to recycling centers depending on the chemical composition or gets directly transported to industrial landfills. However, akin to the case of recycled

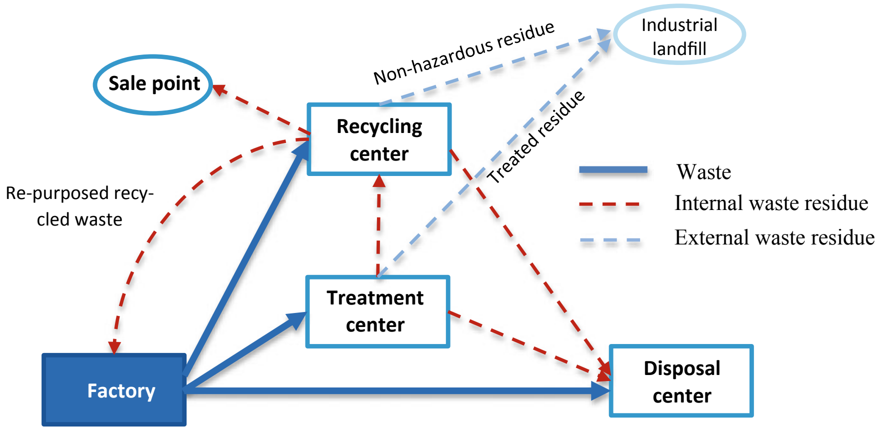


Fig. 1 The proposed hazardous waste transportation network

waste, untreatable waste gets transported to hazardous waste disposal centers for its final disposal. Lastly, hazardous waste that is deemed as non-recyclable and non-treatable gets directly transported to a hazardous waste disposal center, where it is either stored in special containers or incinerated.

3.2 Mathematical Model

The proposed location-routing problem decides: (a) The number and locations of each type of waste facilities (b) The quantity of waste transported through different routes in the network. The main objective of the proposed model is to maximize the total profit of the network, with conditions focusing on limiting the transportation risk in terms of population exposure. the assumptions of the proposed model are as follows:

- The problem is assumed to be a truck transportation problem.
- The amount of the generated waste is known and deterministic.
- The percentages of the three types of generated hazardous waste are known and fixed.
- There are no capacity constraints for roads and vehicles, however, there are capacity constraints for waste facilities.
- Transportation costs are based on the length of routes and fuel prices.
- Every node can be a candidate location for all three types of waste facilities at the same time.
- Population density is assumed to be non-uniformly distributed along a transportation link depending on a maximum set distance from the nearest city center.

- There are budgetary constraints concerning the investment cost for establishing waste facilities.

Model formulation

The notations of the proposed mathematical model are shown in Table 1. The model formulation is as follows:

$$Maxf(x) = \left(\sum_{i\in S} \sum_{j\in J} \sum_{s\in S} r_s * q_{ij}^t + \sum_{i\in R} \sum_{r\in R} m_r * v * x_i \right) - \left(\sum_{i\in I} \sum_{j\in J} c_{ij}^t * d_{ij} * q_{ij}^t + \sum_{i\in I\cup J} c_i^o * q_i^n \right) \quad (1)$$

s.t.

$$\sum_{i\in I} \sum_{j\in J} u_{ij} * d_{ij} * q_{ij}^t * \leq \sum_{i\in I} \sum_{j\in J} \lambda_{ij} * g_s * d_{ij} * u_{ij}, \forall s \in S \quad (2)$$

$$\sum_{j} q_{ij}^t = g_s, \forall s \in S \quad (3)$$

$$q_i^n = \sum_{j} q_{ij}^t, \forall i \in I \quad (4)$$

$$q_i^n = \sum_{j\in R} q_{ij}^t + \sum_{j\in T} q_{ij}^t + \sum_{j\in D} q_{ij}^t, \forall i \in S \quad (5)$$

$$q_i^n = \sum_{j\in D} q_{ij}^t, \forall i \in R \quad (6)$$

$$q_i^n = \sum_{j\in D} q_{ij}^t + \sum_{j\in R} q_{ij}^t, \forall i \in T \quad (7)$$

$$\sum_{i\in I\cup J} c_i^c * x_i \leq \zeta_i, \forall i \in I \cup J \quad (8)$$

$$q_i^n \leq p_i^x x_{i,}, \forall i \in I \quad (9)$$

$$q_i^n \geq p_r^m x_i, i \in I, \forall r \in R \quad (10)$$

$$q_{ij}^t \geq 0, \forall i \in I, \forall j \in J \quad (11)$$

$$q_i^n \geq 0, \forall i \in I \quad (12)$$

Table 1 Mathematical model notations

Notation	Definition
Sets	
I	Set of origin nodes (1,..., i)
J	Set of destination nodes (1,..., j)
S	Set of factory nodes (1,..., s)
T	Set of treatment center nodes (1,..., t)
R	Set of recycling center nodes (1,..., r)
D	Set of disposal center nodes (1,..., d)
Parameters	
g_s	Quantity of generated hazardous waste at factory (s)
c^t_{ij}	Unit transportation cost from origin node (i) to destination node (j)
c^o_i	Unit operation cost at node (i)
c^c_i	Annual investment cost of a facility at node (i)
u_{ij}	Population density per km^2 from origin node (i) to destination node (j)
p^{m_r}	Minimum amount of waste required to establish a recycling facility at node (r)
p^{x_i}	Maximum capacity of a waste facility at node (i)
v	Unit revenue from selling recycled hazardous waste
m_r	Amount of waste transported to recycling facility (r)
r_s	Annual revenue of a factory (s)
d_{ij}	Distance on route (i, j)
ζ_i	Budget for opening a waste facility at node (i)
λ_{ij}	Level of maximum allowed risk on route (i, j)
Decision variables	
q^t_{ij}	Quantity of transported waste from origin node (i) to destination node (j)
q^n_i	Quantity of waste processed at a facility at node (i)
x_i	Binary variable, 1 if a facility is established at node (i), 0 otherwise

$$x_i \in \{0, 1\}, \forall i \in I \tag{13}$$

Objective function (1) maximizes the total profit of the network. Constraint (2) controls the transportation risk. Constraint (3) ensures the flow balance of transported waste and that all generated waste is successfully transported. Constraint (4) ensures that all transportation routes start from a valid source node i and end at a valid destination node j. Constraints (5–7) indicate the allowed directions of transported waste.

Constraint (8) is a budgetary constraint. Constraints (9–10) are capacity constraints, and constraints (11–13) are positivity and binary variable constraints.

4 Computational Results

The model was solved using Python via Gurobi 8.1.1 on Intel Core i5-8250U 1.60 GHz CPU and 8 GB RAM computer. Due to the NP-hard nature of LRP [11] a metaheuristic genetic algorithm was developed to solve different sizes of the problem.

The parameter λ_{ij} (level of maximum allowed risk) is integrated into the problem, which is represented as a percentage of the amount of transported waste. This is done to focus on controlling the transportation risk under a certain level. This value is set to 0.3 based on the conducted sensitivity analysis and the average value used in previous literature.

A non-dominated sorting genetic algorithm II (NSGA-II) is developed. This approach was chosen since it has been proven to be superior in solving LRP models [9]. Moreover, due to the complicated nature of the decision variables in our model, NSGA-II was chosen due to its flexibility in integrating other solution methods. For our problem, a combinational approach was developed where NSGA-II decides the binary decision variables, then in the evaluation part of the algorithm we solve a linear programming model to get the values of the continuous decision variables. The base algorithm was adopted from Pymoo library [12]. An overview of the implemented algorithm is shown in Fig. 2. Multiple experiments were conducted as shown in Table 2. The values of the parameters used are: (1) Half uniform crossover with 0.5 rate (2) Bit-flip mutation with 0.02 rate (3) Population size of 100 (4) 1500 generations.

As observed from the results, the GAP percentages indicate the superiority of the algorithm's performance, especially as the size of the problem increases. In instances larger than 40 facilities, Gurobi solver was not able to achieve a good solution within the specified one-hour run time, while NSGA-II could get remarkably better results in less than half of the run time. With an average GAP of 5.7% and in accordance with

Step 1: Set parameter values according to the previously tested parameter settings; population size, number of generations, crossover and mutation rate, termination criteria.
Step 2: Generate the initial population P_0 with size N by random sampling.
Step 3: Assign ranking to the initial population and sort it into fronts according to non-dominated sorting.
Step 4: Generate offspring population (O_0) by utilizing operators such as mating selection, half uniform crossover, and mutation.
Step 5: Combine parent and offspring populations $T_p = P_o \cup O_0$. Identify fronts F_i, calculate the crowding distances and eliminate duplicate chromosomes.
Step 6: Sort the resulting population again into fronts using non-dominated sorting and calculate the crowding distance for each solution.
Step 7: Solve the linear programming model for each solution to get the values of the continuous decision variables.
Step 8: $P_{i+1} = 0$ and i = 1, while $|P_{i+1}| + |F_i| \leq N$:

$$P_{i+1} = P_{i+1} \cup F_i$$
$$i = i + 1$$

Step 9: Sort solutions in F_i based on the crowding distances.
Repeat steps 4-9 until the termination criteria is met.

Fig. 2 Overview of the implemented algorithm

Table 2 Computational results via Gurobi optimizer and NSGA-II algorithm

Number of facilities		Gurobi results			NSGA-II results		
	(T, R, D)	f	GAP (%)	CPU time (s)	f	GAP (%)	CPU time (s)
5	(1, 2, 2)	305.0	0	18.9	305.0	0	6.2
10	(2, 4, 4)	375.2	0	58.5	375.2	0	27.8
20	(5, 8, 7)	396.2	0	560.8	396.2	0	151.5
30	(9, 11, 10)	521.3	0	1136.3	521.3	0	329.2
40	(11, 15, 14)	2519.0	2.4	3600.0	2567.2	0.4	925.1
41	(11, 16, 14)	3714.8	7.1	3600.0	3914.9	1.6	932.8
42	(12, 16, 14)	4937.4	15.3	3600.0	5498.6	3.5	958.5
43	(12, 16, 15)	8111.2	38.3	3600.0	9828.3	14.1	988.2
44	(12, 17, 15)	9336.2	48.0	3600.0	12,276.6	12.5	1022.2
45	(13, 17, 15)	10,317.1	67.8	3600.0	14,985.6	15.5	1034.3
46	(13, 17, 16)	11,745.2	77.6	3600.0	17,994.1	15.9	1148.9
47	(13, 18, 16)	NA (not feasible)			19,756.2	NA	1274.5

* Number of factories for all instances is set as 4

the complexity of the model, the proposed algorithm has proved to yield satisfactory results for medium and large size problems.

Case study: Jordan

In order to investigate the solution for a large size problem, the model was applied to the waste management network in the North and Central regions of Jordan. Those regions were chosen due to the prominent presence of textile factories. According to the Netherlands Enterprise Agency [13] the textile sector in Jordan has grown by 35% in the span of five years between 2012 and 2017. However, the generated waste remains disposed of in an unmanaged manner.

The candidate locations were chosen according to the locations of pre-existing waste collection centers, as well as some additional locations in non-densely populated areas. The necessary information including the population densities, distances, and locations of existing facilities were taken from the report published by Japan International Cooperation Agency in 2016 [14]. Data on annual profits of textile

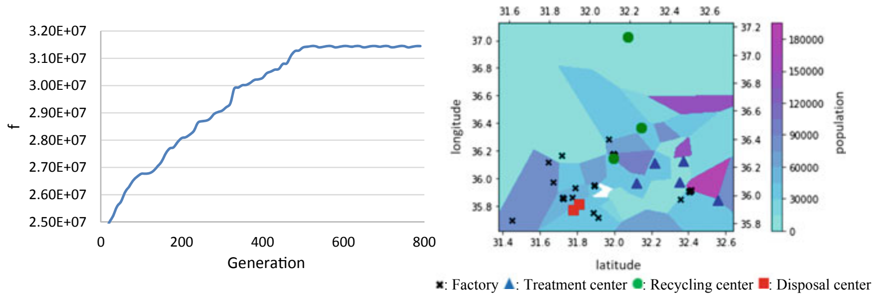


Fig. 3 **(a)** Convergence of the algorithm **(b)** Map of the chosen locations

factories were taken from Jordan textile incoming trade mission to the Netherlands report published in 2018 [15].

The problem was solved with a dataset of 54 factories and 18 candidate locations of waste facilities via the developed NSGA-II algorithm. Figure 3a illustrates the convergence curve of the algorithm, based on the best fitness value of each generation. The algorithm stabilizes after around 550 generations, noting that it was terminated after one hour of run time. A map showing the chosen locations of waste facilities and the corresponding population densities is shown in Fig. 3b.

The best solution found in terms of profit is around 31 million EUR annually. This value seems reasonable as a combined profit of the network, since most factories are small and medium size with an annual profit of less than 1–2 million EUR. Moreover, the chosen locations show the trade-off between the profit and potential transportation risk, where some locations were chosen in relatively densely populated areas due to the larger profit resulting from lower transportation cost. Nonetheless, it is worthy to mention that the highest populated areas shown in purple in Fig. 3b have been avoided while still choosing locations as close as possible to where most factories are located.

5 Conclusions

In this study, we proposed a profit-oriented mixed-integer linear programming model for the hazardous waste location routing problem. The main contribution of this research is to offer a new lens on how to consider the overall profit in hazmat transportation networks. The mathematical model was solved by Gurobi solver and NSGA-II algorithm due to its NP-hard nature. The computational results confirm the superiority of the proposed algorithm in solving medium and large size problems within an acceptable time. The model was then applied to real life case study in Jordan and yielded reasonable and applicable results.

References

1. Aydemir-Karadag, A.: A profit-oriented mathematical model for hazardous waste locating-routing problem. J. Cleaner Prod. **202**, 213–225 (2018). https://doi.org/10.1016/j.jclepro.2018.08.106
2. Glickman, T.S.: Rerouting railroad shipments of hazardous materials to avoid populated areas. Accid. Anal. Prev. **15**(5), 329–335 (1983)
3. Holeczek, N.: Hazardous materials truck transportation problems: a classification and state of the art literature review. Transp. Res. Part D **69**, 305–328 (2019). https://doi.org/10.1016/j.trd.2019.02.010
4. Zografos, K., Samara, S.: Combined location-routing model for hazardous waste transportation and disposal. Transp. Res. Rec. **1245**, 52–59 (1989)
5. Wang, H., Du, L., Ma, S.: Multi-objective open location-routing model with split delivery for optimized relief distribution in post-earthquake. Transp. Res. Part E **69**, 160–179 (2014). https://doi.org/10.1016/j.tre.2014.06.006
6. Beneventti, G.D., Andres, B., German, P., Vladimir, M.: A multi-product maximin hazmat routing-location problem with multiple origin-destination pairs. J. Cleaner Prod. **240**, 118193 (2019). https://doi.org/10.1016/j.jclepro.2019.118193
7. Iris Abril, M., Julian, M., Francisco, M., Trinidad, G., Rafael, C.: Solving a bi-objective transportation location routing problem by metaheuristic algorithms. Eur. J. Oper. Res. **234**, 25–36 (2014). https://doi.org/10.1016/j.ejor.2013.09.008
8. Pichka, K., Bajgiran, H., Petering, H., Jang, J., Yue, X.: The two echelon open location routing problem: mathematical model and hybrid heuristic. Comput. Ind. Eng. **121**, 97–112 (2018). https://doi.org/10.1016/j.cie.2018.05.010
9. Nedjati, A., Gokhan, I., Jamal, A.: Bi-objective covering tour location routing problem with replenishment at intermediate depots: formulation and meta-heuristics. Comput. Industr. Eng. **110**, 191–206 (2017). https://doi.org/10.1016/j.cie.2017.06.004
10. Boyer, O., Hong, T., Pedram, A., Yusuff, R., Zulkifli, N.: A mathematical model for the industrial hazardous waste location-routing problem. J. Appl. Math. **2013**, 435272 (2013). https://doi.org/10.1155/2013/435272
11. Zhong, S., Cheng, R., Jiang, Y., Wang, Z., Larsen, A., Nielsen, O.: Risk-averse optimization of disaster relief facility location and vehicle routing under stochastic demand. Transp. Res. Part E **141**, 102015 (2020). https://doi.org/10.1016/j.tre.2020.102015
12. Blank, J., Deb, K.: Pymoo: multi-objective optimization in python. IEEE Access **8**, 89497–89509 (2020). https://doi.org/10.1109/ACCESS.2020.2990567
13. Netherlands Enterprise Agency: Value Chain Analysis (VCA) Garments, Jordan (2019). Publication number: RVO-027–1901/CBI-RP
14. Japan International Cooperation Agency (JICA), Kokusai Kogyo Co., Ltd.: Data collection survey on waste management in northern region accepting Syrian refugees in The Hashemite Kingdom of Jordan. 7R-JR-16–010 (2016)
15. Nederlandse Export Combinatie: Jordan textile incoming trade mission to the Netherlands, March 18–23, Textile Companies (2018)

A Bi-Objective Integer Linear Optimization Model for Post-Departure Aircraft Rerouting Problem

Miriam F. Bongo and Charlle L. Sy

Abstract In as early as the 1980s, air traffic flow management actions (ATFM), as supplementary strategies to match the demand for air travel with the available resource capacities, have been widely discussed and evaluated based on its implementation and probable trade-offs between conflicting and diverse interests of stakeholders in the commercial aviation industry. Among the ATFM actions—ground holding, airborne holding, speed controlling, and rerouting—rerouting is found to be a viable recourse particularly when flights are already at its en-route phase, where the presumed and more favored based on safety considerations, holding of flights on the ground, becomes completely infeasible. Some research works put forward relevant solution approaches including deterministic and stochastic mathematical programming models, machine learning algorithms, and simulation models. Despite the relevance and validity demonstrated by such models in testbed environments, even on a large-scale basis, these models failed to sufficiently capture the individual and collective interests of stakeholders altogether. Considering that the decision process in the air transportation system is taken part by stakeholders (i.e., airlines, air traffic control), previous research works tend to satisfy only one stakeholder by incorporating one or more of its interests (e.g., cost minimization, reduction of distance traveled). Such a case does not take full regard to how a stakeholder-specific solution might affect another stakeholder's preference. Therefore, this paper aims to address the post-departure aircraft rerouting problem by proposing a multiple stakeholder-based target-oriented robust-optimization (MS-TORO) approach that incorporates the individual interests of stakeholders. A hypothetical case study is conducted to

M. F. Bongo (✉) · C. L. Sy
Industrial and Systems Engineering Department, Gokongwei College of Engineering, De La Salle University, Manila, Philippines
e-mail: miriam_bongo@dlsu.edu.ph

C. L. Sy
e-mail: charlle.sy@dlsu.edu.ph

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_43

illustrate the proposed model. It can be noted that a significant shift of route preference occurs as goals are aligned in terms of the individual interests of the stakeholders and that of their collective goal. The results of this work can provide practical insights to stakeholders in the course of decision-making in a particular area of the air transportation domain.

Keywords Airline · Air traffic management · Aircraft rerouting · Multi-objective optimization · Safety

1 Aircraft Rerouting Problem

As soon as the aircraft flies past the departure fix and is at the en-route airspace, it becomes a part of the system under evaluation as in Fig. 1. Suppose that at this point a disruptive event has occurred and has significantly affected the capacity of the original route established in the flight plan. Specifically, the capacity of the route is reduced, if not totally unavailable for utilization. Therefore, the set of aircraft at the given period awaiting rerouting clearance is prompted to take on an alternate route. For simplicity purposes, the set of aircraft considered is assumed to belong to the same air carrier, thus, maintaining the consistency of decision as may be necessary for the rerouting process. For an alternate route to be selected, it has to satisfy the preferences of multiple decision-makers. These decision-makers are represented by stakeholders of the airport management, airline industry, and air traffic management [1]. Each stakeholder has an individual auxiliary interest they tend to achieve; collectively, a common goal is prioritized by these stakeholders. The airport management seeks to maximize the utilization of resources, specifically the airspace in this case [2]. On the other hand, the airline industry aims to achieve the minimum cost associated with rerouting flights to an alternate route [3]. Lastly, air traffic management pursues an expeditious and orderly flow of air traffic along with the top priority of maintaining the safety of flight operations [4].

While these stakeholders have endeavored to uphold their individual auxiliary interests, the interest of maintaining a safe operation of flights in all phases, especially the critical ones, is attended to by the stakeholders with utmost priority among others [4]. Since safety in the context of the air transportation domain is relatively difficult to quantify [5], a more suitable indicator during the en-route phase to establish safety can be represented by maintaining a minimum separation distance between aircraft to prevent conflicts among aircraft [6]. The minimum separation distance between aircraft is patterned after a prescribed standard established by the ICAO [7]. Generally, an alternate route is considered to be safer when it can potentially prevent conflicts by providing more separation distance between flights while accommodating all flights at a given period. During a disruptive event, the capacity of routes is dramatically reduced, if not restricted, thereby becoming the main cause of congestion [8]. Such an effect poses a major threat to the stakeholders particularly

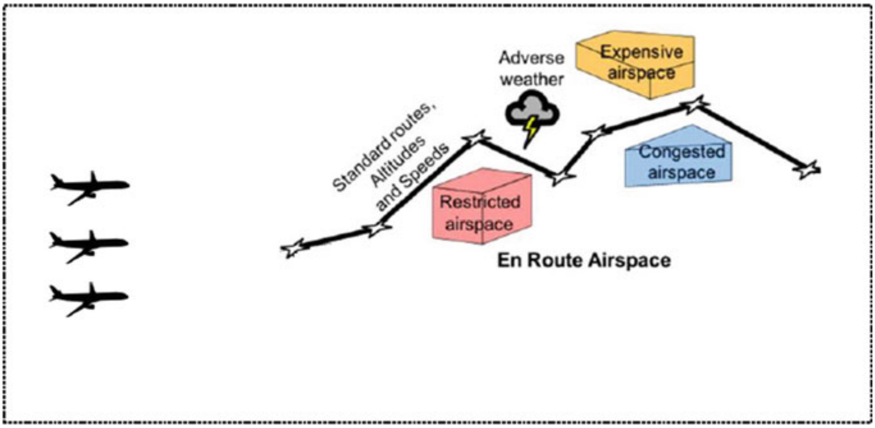


Fig. 1 System of an en-route flight subject to a rerouting procedure (Adapted from Reynolds, 2014)

since disruptions caused by adverse weather, for instance, calls for a tactical planning scheme. That is, implementing a course of action on the day of operations. As disruptions occur at the tactical planning level, there exists uncertainty in capacity of routes. This phenomenon, together with the cost implications of rerouting an aircraft at its en-route phase, forms part of the decision-making process endeavored by stakeholders to address with urgency.

Case-specific, the aircraft rerouting problem, although is widely tackled in the literature and has given rise to significant solutions to address one or more of its problem areas, still presents an immense avenue for further improvement due to the shortcomings on how prior decision processes are framed as well as the solution approaches formulated. To advance the body of knowledge in aircraft rerouting, it becomes a necessity to shape the issue focusing on aircraft post-departure phase with a collaborative decision-making process involving the common goal of ensuring safe operation of flights and the individual auxiliary preferences of each stakeholder concerned (e.g., airport management, airline industry, air traffic management). Among other goals, the safety of aircraft operations as represented by the minimum separation distance between aircraft traversing around the same path specifically needs to be set as an objective function given that the topmost priority of stakeholders points towards ensuring safe operations. Also, in recognition of the dynamic and unpredictable variations of resource capacities during disruptions, the uncertainty of resources, specifically routes in airspaces, is deemed appropriate to be treated as a target-oriented robust optimization problem.

Previous research works viewed the rerouting problem through the lens of only one stakeholder (e.g., airline industry), thus, formulated solution approaches according to their individual goals (e.g., minimizing total delay costs) [9, 10]. Setting the rerouting problem in such a way neglects the preferences of other stakeholders (i.e., airport management, air traffic management) involved in the decision process as well as the general goal of the air transportation domain. It is essential to satisfactorily cover the system-wide preference of all stakeholders in order to arrive at a globally acceptable solution. Thus, setting the solution approach with respect to safety in aircraft operations while taking into consideration the preference of stakeholders as system constraints enables a more comprehensive and all-encompassing solution satisfying not only one stakeholder but all that is involved.

At a closer level, the rerouting process can be a rather complex and rigorous decision problem that requires a huge sense of urgency especially when the need for rerouting occurs at the post-departure phase of a flight. Stakeholders, as decision-makers, tend to frame solutions that satisfy their own interests and present such solutions to other stakeholders for probable implementation. For instance, when disruptions occur and require immediate rerouting decisions, airline companies typically request a rerouting to air traffic controllers which might be disagreeable to both entities due to conflicting preferences. While pilots prefer a shorter alternate route that can supposedly allow a flight to traverse from one sector to another at a minimum flight time, air traffic controllers may not issue a clearance to pilots to take on that route due to potential traffic issues at hand [11]. On the other hand, a longer alternate route that provides more room for controllability and conflict avoidance may appear favorable to air traffic controllers but is considered otherwise for pilots who consider the economy of aircraft operations. Therefore, framing the rerouting decision process according to the individual auxiliary interests of stakeholders as well as their collective goal can provide a solution that minimizes the penalization of stakeholders while ensuring that the common goal is satisfactorily achieved. Such a manner enforces a collaborative effort between stakeholders which allows for the individual auxiliary interests to be incorporated in the decision-making process while sustaining the fulfillment of their common goal.

As disruptions occur during regular aircraft operations, the uncertainty of resource capacities becomes evident which prompts decision-makers to take on measures that shall impose the least impact on the entire network of flights. Unfortunately, the presence of uncertainty in the general air transportation domain, specifically in rerouting, has been treated as stochastic in nature notwithstanding that the probability density functions associated with such uncertain parameters may be difficult to obtain at the same time quantify [1, 12].

Therefore, it is necessary to design a modeling framework that embeds the theoretical and mathematical foundations of the stakeholder theory and the classical TORO model, respectively, for the aircraft rerouting problem. This framework must be developed to ascertain that the interests of multiple stakeholders are balanced accordingly via multiple targets at the same time deal with uncertainties in an efficient manner. The proposed multiple stakeholder-based TORO approach closely follows the mathematical properties of the classical TORO approach including the elimination of specific assumptions on the probability density functions of uncertain parameters. However, unlike the classical TORO approach, the proposed framework will allow decision-makers to specify different targets and a different set of solutions that satisfy each target. That is, the different risk profiles of multiple stakeholders can be distinguished along with the corresponding solutions that suit their profiles. While a different set of solutions can be generated based on the risk profiles of multiple stakeholders, an overall solution can be obtained by compromise of solutions.

2 A Bi-Objective Collaborative Model

A novel multiple stakeholder-based target-oriented robust optimization approach (MS-TORO) model is developed based on the need for a framework that shall balance the interests of multiple stakeholders with inherent diverse, often conflicting, interests. The proposed model also integrates the mathematical properties of the classical TORO model developed by Ng and Sy [13] by extending the satisficing solution to an index that represents the degree of fulfillment to each stakeholder's interest. Also, the proposed model can be applied to systems that involve the participation of multiple stakeholders in the decision-making and the evidence of uncertainty as well. Some examples of such systems include the supply chain, logistics, enterprise management, and air transportation domain.

As a case application, the proposed MS-TORO model is formulated to address the post-departure rerouting problem which satisfactorily considers key points in the decision process such as the collaborative decision making among stakeholders, the incorporation of safety as a focal objective of the model proposed, and the consideration of uncertainty in resource capacities. To the best of the author's knowledge, this work is the first to propose a new framework that addresses a pressing issue in the air transportation domain. With the application of the proposed MS-TORO model, system targets can be carefully accounted for along with the inherent uncertainties present in the system characteristics of this domain. This methodological advancement is a huge leap from the traditional approaches previously presented in the literature which only focuses on either the deterministic version of the air traffic flow management problem or its stochastic version which still apparently renders inadequate due to how it handles uncertainty in the system. Therefore, stakeholders may be able to effectively frame policies on the rerouting problem with the aid of the proposed model given that the level of uncertainty can be clearly distributed established all throughout the decision planning horizon.

2.1 Model Formulation

Consider a set of flights scheduled to depart to two different destination points. At the en-route phase, the flights are assumed to experience air traffic congestion which prompts the stakeholders to implement a rerouting procedure. Suppose that there exist three independent alternate routes for each destination point. These alternate routes vary according to its length and costs associated with it. Suppose further that two stakeholders—airline industry and air traffic management—need to collaborate in the most immediate manner to arrive at an agreeable solution to the rerouting procedure.

Typically, the rerouting process can also be participated by another stakeholder, the airport management, but for the case of this paper, it is assumed that this stakeholder is indifferent to whatever alternate route may be chosen with respect to the interests of the two other stakeholders.

Note further that the airline industry aims to minimize its operating cost while performing the rerouting procedure. On the other hand, air traffic management is concerned with minimizing the probable delays to be accumulated should a longer alternate route is chosen. The proposed MS-TORO model aims to support a choice problem of selecting a candidate alternate route for the scheduled flights to take. Therefore, the proposed model is patterned after an integer linear programming model with the following model components:

Decision variables. For the case of this paper, we refer to the formulations presented in the research work of Bongo and Sy [14] with a further reference to Bertsimas and Stock-Patterson [15]. The decision variables represent the set of alternate routes that flights may traverse to reach its destination point. Note that the rerouting procedure at the en-route phase is considered for this case, therefore, no implementation of other air traffic flow management actions (e.g., excess ground holding, airborne holding, or speed controlling) is assumed to be taken. Note further that the rerouting process is only intended to search for another feasible route to reach the intended destination point; as such, no diversion of flights is also regarded. Specifically, the decision variable involved in the proposed rerouting model is represented as follows:

$$w_j^d = \begin{cases} 1, & \text{if a flight going to destination } d \text{ selects route } j \\ 0, & \text{otherwise} \end{cases}$$

Notation: For the model formulation, the following notations are used:

- J Set of routes,
- D Set of destination points,
- f Set of flights,
- c Cost of using alternate route j,
- k Stakeholders,
- τ_{k} Performance target of stakeholder k,
- R Capacity of alternate route j.

Objective function: In the event of a post-departure aircraft rerouting problem, the airline industry and air traffic management each uphold a goal to satisfy their own interests. For the first objective function, the airline industry seeks to minimize operating costs composed of using an alternate route. For the second objective function, the air traffic management aims to minimize the delays incurred (i.e., in time units) in traversing an alternate route that is relatively longer than the initial route established. The two objective functions are rather conflicting as satisfying the first objective may imply the selection of an alternate route that is not necessarily favorable to the second objective. Therefore, to reconcile this difference in perspective, the proposed model introduces a metric to represent the deviation of performance targets of stakeholders. Such a metric is designed to arrive at a solution that is agreeable between stakeholders to a certain degree. As such, the deviation metric, θ, is minimized. To put this into context, consider the following objective function as in Eq. (1):

$$\min \theta \tag{1}$$

Constraints: To consider the bi-objective model with respect to the interests of the airline industry and the air traffic management, these objectives are translated as constraints along with the following system constraints as follows:

$$\sum_{j=1}^{n} \left(c_j^d \cdot w_j^d\right) \leq \tau_k (1+\theta) \quad \forall d \in D \tag{2}$$

$$\sum_{j=1}^{n} \left(r_j^d \cdot w_j^d\right) \leq \tau_k (1+\theta) \quad \forall d \in D \tag{3}$$

$$\sum_{j=1}^{n} \left(f \cdot w_j^d\right) \leq R \quad \forall j \in J, \forall d \in D \tag{4}$$

$$\sum_{j=1}^{n} w_j^d = 1 \quad \forall j \in J \tag{5}$$

$$w_j^d \in \{0, 1\} \tag{6}$$

$$0 \leq \theta \leq 1 \tag{7}$$

Constraints (2–3) represent the objective functions of the stakeholders each translated into a constraint. In specific, constraint (2) accounts for the objective of the airline industry to minimize operating costs composed of the cost of using a flight route. This constraint should be lesser than or equal to the deviated performance target as represented by $\tau_k(1+\theta)$. .. On the other hand, constraint (3) represents the objective of the air traffic management to minimize total delays incurred due to the rerouting procedure. This constraint should also be lesser than or equal to the performance target of the air traffic management and its corresponding deviation as in $\tau_k(1+\theta)$. The performance targets of stakeholders correspond to their target budget and target delays, respectively. Constraint (4) enforces that the number of flights considered should not exceed the capacity of an alternate route in terms of number of aircraft. Constraint (5) ensures that only one alternate route is selected. As for constraint (6), this defines that the decision variable w_j^d is a binary variable. Lastly, constraint (6) specifies that the deviation metric, θ, can only take on a numerical value within the closed interval 0 to 1.

3 Results and Discussion

Consider a hypothetical scenario of post-departure aircraft rerouting problem as follows. Suppose there are 30 flights en-route to two particular destination points (i.e., 15 flights scheduled for each destination). Suppose further that an air traffic congestion is noted during the en-route phase which calls for the implementation of a rerouting procedure. For both destination points, there are three alternate routes with various characteristics in cost, length, and capacity (see Table 1). Assuming that the ground resources such as slots and gate assignments will be made available once the rerouted flights reach their respective destination points, the airline industry and air traffic management then have to confer and decide on an alternate route that will satisfy both interests. Note that the airline industry has set a performance target of $200 as budget allocation while the air traffic management has set a performance target of 25 min in terms of total delay incurred due to the rerouting procedure. The proposed model with the parameters set is run in a Demo Lingo/Win64 software with results shown in Table 2.

Implementing the proposed model provides a choice of an alternate route for each destination point that is agreeable between the two stakeholders involved. Recall that the airline industry and air traffic management apparently have a conflicting interest as, for instance, with respect to cost implications, the airline industry would typically select the least expensive route (e.g., route A and route D for destination 1 and destination 2, respectively). On the other hand, the air traffic management would

Table 1 Parameters of the proposed model

Destination	Alternate route	Cost (in $)	Length (in NM)	Delays (in minutes)	Capacity (in no. of aircraft)
1	A	329	100	40	20
	B	500	90	30	18
	C	400	75	15	15
2	D	100	75	25	15
	E	300	85	35	17
	F	500	75	25	15

Table 2 Computational results

Destination	Alternate route	Results	Deviation metric
1	A	1	0.645
	B	0	
	C	0	
2	D	1	
	E	0	
	F	0	

favor an alternate route that will result in the least delays in time units (e.g., route C and route D or F for destination 1 and destination 2, respectively).

With the proposed MS-TORO model, a compromise solution can be obtained as in Table 2 showing that alternate route A and route D for destination 1 and destination 2, respectively, should be selected to satisfice both stakeholders' interests. Such compromise can be made possible when the deviation metric is observed. That is, by selecting route A and route D, the stakeholders should be able to adjust its performance targets to 64.5% such that a solution that satisfices its own interest and another may be generated. The deviation metric further implies that (1) the airline industry should extend its performance targets (i.e., budget allocation) to as much as $329 and (2) the air traffic management should also adjust its delay allowance to as high as 41 min. This is so that for an agreeable solution to be obtained, corresponding adjustments to the performance targets should be on the part of both stakeholders involved.

4 Concluding Remarks

This paper presents a novel approach to simultaneously consider the conflicting interests of stakeholders. This approach is termed multiple stakeholder-based target-oriented robust-optimization (MS-TORO). The proposed model aims to generate a solution that can satisfice the interests of stakeholders by introducing a deviation metric that signifies the extent to which adjustments on the performance targets should be made. A hypothetical case scenario in a post-departure aircraft rerouting problem is presented under two distinct destination points. The results primarily show that despite the initial conflict in the interests of stakeholders, the proposed model is able to develop compromise solutions that can simultaneously consider various interests.

Such a contribution to the literature is deemed significant as real-world applications do involve the participation of decision-makers with various, often conflicting, interests. Therefore, a solution approach that can handle such a case need to be developed; one that is computationally tractable and does not require considerably long computational times.

References

1. Chang, Y.H., Solak, S., Clarke, J.P.B., Johnson, E.L.: Models for single-sector stochastic air traffic flow management under reduced airspace capacity. J. Oper. Res. Soc. (2015). https://doi.org/10.1057/jors.2015.53
2. Janić, M.: Reprint of "Modelling the resilience, friability and costs of an air transport network affected by a large-scale disruptive event." Transp. Res. Part A Policy Pract. **81**, 77–92 (2015)
3. Zhang, D., Lau, H.H., Yu, C.: A two stage heuristic algorithm for the integrated aircraft and crew schedule recovery problems. Comput. Ind. Eng. **87**, 436–453 (2015)

4. Tobaruela, G., Fransen, P., Schuster, W., Ochieng, W.Y., Majumdar, A.: Air traffic predictability framework development, performance evaluation and application. J. Air Transp. Manag. **39**, 48–58 (2014)
5. Kalemba, N., Campa-Planas, F.: Safety and the economic and financial performance in the airline industry: is there any relationship? Aviation **23**(1), 7–14 (2019)
6. Novak, A., Havel, K., Adamko, P.: Number of conflicts at the route intersection—minimum distance model. Aviation **23**(1), 1–6 (2019)
7. ICAO.: Air traffic management (Doc 4444, Code 7700, 16th ed.). Retrieved from https://ops.group/blog/wp-content/uploads/2017/03/ICAO-Doc4444-Pans-Atm-16thEdition-2016-OPSGROUP.pdf. Retrieved on 26 Aug 2020 (2016)
8. Kammoun, M.A., Rezg, N.: An efficient hybrid approach for resolving the aircraft routing and rescheduling problem. J. Air Transp. Manag. **71**, 73–87 (2018)
9. Tompa, R.E., Kochenderfer, M.J.: Efficient aircraft rerouting during commercial space launches. New Space **7**(1), 12–18 (2019)
10. Srivastava, A.: Generating realistic reroutes to assess air traffic impact of blocked airspaces. In: 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC), pp. 1–9. San Diego, CA, USA (2019). https://doi.org/10.1109/DASC43569.2019.9081707
11. Aktürk, M.S., Atamtürk, A., Gürel, S.: Aircraft rescheduling with cruise speed control. Oper. Res. **62**(4), 829–845 (2014)
12. Agustín, A., Alonso-Ayuso, A., Escudero, L.F., Pizzaro C.: On air traffic flow management with rerouting. Part II: Stochastic case. Eur. J. Oper. Res. **219** 167–177, (2012)
13. Ng, T.S., Sy, C.L.: An affine adjustable robust model for generation and transmission network planning. Electr. Power Energy Syst. **60**, 141–152 (2014)
14. Bongo, M.F., Sy, C.L.: An integer linear programming formulation for post-departure air traffic flow management. ASEAN Eng. J. **11**(2), 101–117 (2021)
15. Bertsimas, D., Stock Patterson, S.: The air traffic flow management problem with enroute capacities. Oper. Res. **46**(3), 406–422 (1998)

Forecasting the Economic Number of Times Sheet Metal Subject to Corrosion Should Be Painted Before Replacement

R. Assis and P. Carmona Marques

Abstract The present paper describes a method for forecasting the economic number of times a sheet metal, integrated in a structure, and subjected to high corrosion, must be repainted before its thickness reaches a minimum limit, advising for its replacement. Within this approach, one is acknowledged of all the relevant costs originated, predictably, over the life cycle of the sheet metal, demonstrating the utility of combining descriptive functions (of the physical phenomena from the field of materials engineering) with financial analysis' functions, thus completing what one can properly call—technical–economic analysis—whose opportunity is verified so many times in engineering economy and maintenance.

Keywords Economic analysis · Optimization · Forecasting · Decision-making

1 Introduction

There are many cases of dimensioning in engineering in which the decision variable, safeguarding security, and environmental impact aspects, can assume different values (standardized by manufacturers and often described in technical tables). The selection of the most reliable value results, in the last analyses, of the ponderation of several costs, predicted along the structure element's life cycle, considered relevant to the decision support process. This value of the decision variable makes the sum of all the costs minimum (investment, exploration, and deactivation). It is, therefore, the optimum value from the economic perspective, which should ideally be the elected

R. Assis (✉) · P. Carmona Marques
Research Centre in Industrial Engineering, Management and Sustainability, Lusófona University, Lisbon, Portugal
e-mail: p1028@ulusofona.pt

P. Carmona Marques
e-mail: p4803@ulusofona.pt

P. Carmona Marques
ISEL, Polytechnic Institute of Lisbon, Lisbon, Portugal

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_44

from several alternatives. This is, with no doubt, a matter of great importance, when one takes conscience that engineers receive mostly a technological education, and of which, in matters of economic and financial analyses, they are practically out of bounds. Opportunities to introduce economic rationality into decisions, at the level of the project, or at the level of productivity improvements throughout the exploration phase of much equipment, are therefore lost. In fact, it is a widespread and incorrect practice to decide only based on the lowest acquisition price (cost of investment). Being resources always scarce, it is consensual to say that it is important to select the most economical alternative among all viable alternatives (according to other criteria). However, what is many sometimes forgotten (or unknown?) is that being more economical means presenting the smallest cost of the life cycle and not the smallest investment [1–4].

There are structures that work in immerse hostile environments (liquid or gaseous) that degrade with relative velocity. After some time of their service entrance, occurs the first stage of partial recovery of exposed materials, to which others will follow. This recovery is carried out applying surface protection materials (painting, metallization, etc.). After some time, the physical characteristics of some parts of the structure will have degraded to minimum safety limits, making it inevitable to replace some of its components. Steel sheets forming part of the side and the deck of vessels, or of the side walls and tanks for chemical treatment by immersion, are some examples. This type of structure often has a life expectancy of some decades of years, which makes it economically feasible to partially substitute constituent materials for the structure rather than the radical decision for definitive slaughter [5–7].

In this type of situation, each stop for costs of materials, application costs, shipyard occupancy costs and opportunity. The latter are proportional to the immobilization time. On the other hand, degradation is contained for some time, stretching the life of the exposed materials. When the minimum limit of some measurable physical characteristic (thickness) is reached at some point in the structure, the degraded component (sheet metal, profile, etc.) must be replaced immediately. In practice, it will be replaced as soon as the last measurement is known, and further surface recovery is deemed unnecessary. At this time, costs will normally be much higher than a painting operation.

2 Methods

Supposing the case of a sheet metal in those conditions, in which the thickness constitutes the critical parameter. A problem of this nature can then be formulated as follows: once the initial thickness e_0 and the minimum thickness limit e_m are known, what should be, in terms of economic perspective, the ideal number of painting times n of the sheet metal before its thickness reaches values very close to the minimum thickness limit e_m? or in another way, what should be the fraction Δe of the permissible degraded tolerance range of the thickness ($e_0 - e_m$) that minimizes the sum of the cost of replacing the sheet metal (investment) and the cost of various painting operations (maintenance) over the lifetime, i.e., until the minimum limit thickness is reached?

Assuming also known the values: (i) the cost of replacing the sheet metal with a new one C_s (material, disassembly and assembly, painting, shipyard, and opportunity), (ii) the cost of a painting operation Cp (painting, shipyard, and opportunity), (iii) the minimum rate of return $i\%$ that the company has established as policy for this type of investments, (iv) and the time of a painting operation T (periods) before the degradation restarts.

The next step consists in knowing the expected evolution of the thickness of the sheet metal in function of time (degraded function). This evolution is, normally, of the following type [8].

$$e = e_0 - \alpha \times t^{\beta} \tag{1}$$

where: e = thickness of the sheet metal over t periods (weeks, months, years, etc.), e_0 = initial thickness of the sheet metal, α = initial velocity of degradation, t = time (months, years, etc., or generally, periods), and β = acceleration of degradation.

The graphical representation of the past function can be observed in Fig. 1.

However, whenever the plate reaches the thicknesses $(e_0 - \Delta e)$, $(e_0 - 2 \times \Delta e)$, $(e_0 - 3 \times \Delta e)$, …, $(e_0 - n \times \Delta e)$, to values very close to e_m, the sheet metal will be repainted. Therefore, its degradation will be interrupted n times during $n.T$ units of time. Thus, if the restart of the degradation occurs at each point on the curve where it was previously interrupted, that function ceases to be continuous and starts to present levels which at instants 0, t_1, t_2, t_3, …, t_n, each with the duration of T periods, as in Fig. 2.

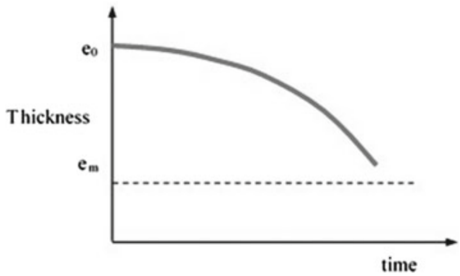


Fig. 1 Time evolution of the thickness of the sheet metal without painting time intervals

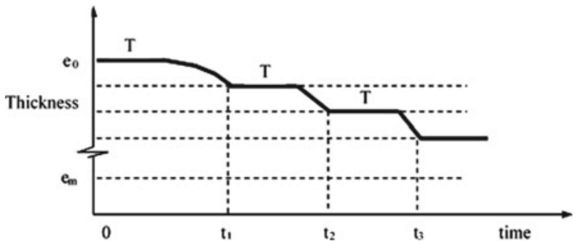


Fig. 2 Time evolution of the thickness of the sheet metal with painting time intervals

This approach consists in calculating, for each decrement of thickness (Δe) and adopting the constant price method, the total cost equals the sum of the investment cost (i.e., the cost of replacing the degraded sheet with a new one) and the maintenance cost (i.e., the cost of painting which will be followed to extend its life). Once the total cost of each alternative is known, the optimum value can be selected, which obviously corresponds to the lowest value of that. As the life span of the sheet metal for each alternative is different, as it varies proportionally with the frequency of the paintings, and assuming the continuity of the life of the equipment of which the sheet metal is an integral part, the appropriate method for comparing costs of each alternative should be that of uniform cost equivalent (or income) [10]. The costs to take into consideration are then [8]:

1. The sum of the costs of painting operations that will take place at times t_1, t_2, t_3, …, t_n, standardized for the time interval between now and the time t_s (instant in which the sheet will be replaced again), equivalent to the sum of N times the cost C_p of a painting operation and updated at the periodic rate i_p%.

$$A_p = \left(A/P; i_p; t_s\right). \sum_{1}^{N} C_p.\left(P/F; i_p; t_n\right) \tag{2}$$

2. The cost of replacing the standard C_s sheet at the periodic rate i_p% for the time interval between now and the time t_s,

$$A_s = C_s.\left(A/P;\mathrm{i}_p;\mathrm{t}_s\right) \tag{3}$$

The cash-flow diagram representing these costs can be seen in Fig. 3. In this, it was assumed that the instant of analysis (today or instant 0) coincides with one of the instants in that the plate has just been replaced and will go into service.

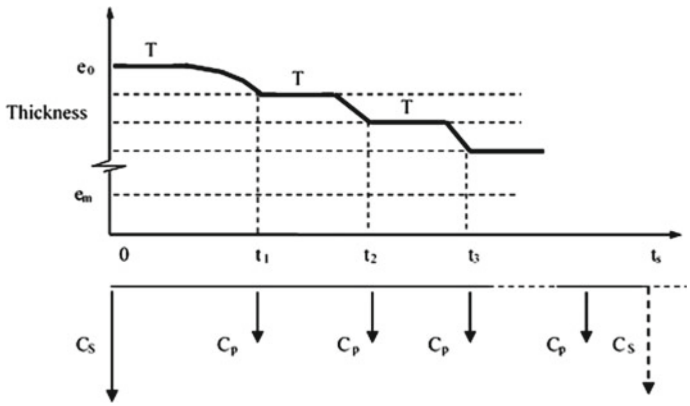


Fig. 3 Cash flow showing the moments of the foreseen painting operations and the sheet metal replacement

Adding up the two costs A_s and A_p for each considered alternative of thickness decrease Δe, the optimum value Δe, the one that corresponds to the lowest total uniform cost, is found. The solution may be obtained by a numerical calculation method.

3 Case Study and Results

Let us assume the case of sheet metals on the deck of a vessel. Surface degradation occurs in a very heterogeneous manner, with areas of higher incidence occurring. In the new state, the sheet has a thickness of 10 mm, and it is specified that no point on its surface should ever have a thickness of less than 6 mm. The registration of this type of vessel and the anticipated conditions of use (or resulting from laboratory), one can estimate the values of the expression parameters of the temporal degradation (in months) of the thickness: $\alpha = 0.2$ and $\beta = 2.0$. The average time during which the paint sufficiently protects the plate, i.e., without any signs of superficial corrosion, is 6 months in accordance with vessel's company. Each painting operation shall be carried out during the operating time of the vessel (at sea or on a shipyard). The estimated costs of each painting and the general replacement of the sheet are, respectively, 10 K€ and 15 K€. The minimum real annual interest rate set by the company for investments in the field of maintenance is 24%. Let us first represent the cash flow diagram if the instant of analysis coincides with the instant when the vessel leaves the yard—in the new state or after a replacement of the sheets—and (re)starts its activity (Fig. 4).

The analysis of this problem does not allow the representation of the result by a continuous function. The result must be found by attempts testing several possible values of the thickness decrease Δe. Let us assume discrete variations of the decrement multiples of 0.5 mm. The thickness of the sheet metal starts at 10 mm and can be reduced up to 6 mm. Assuming the magnitude of the decrement Δe to be 0.5 mm, one gets 4/0.5 = 8 alternatives. Suppose, too, that the degradation function is resumed at the point where it was suspended each time the sheet metal is cleaned and repainted. The costs are then:

Instant t_1 is given by: $t_1 = T + t_{01}$.
And the thickness e_1 is given by: $e_1 = e_0 - \Delta e$.
Given that: $e_1 = e_0 - \alpha \times {t_{01}}^{\beta}$
It comes that: $t_{01} = [(e_0 - e_1) \div \alpha]^{1/\beta}$
Therefore: $e_1 = 10 - 0.5 = 9.5$ mm.

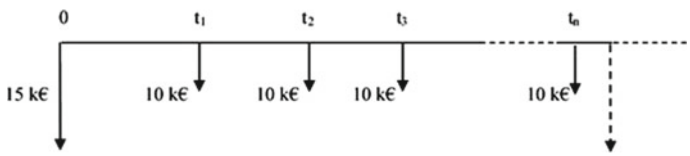


Fig. 4 Planned cash flow

Table 1 Sheet metal thickness reduction with time

Thickness (mm)	Month
10	0
9.5	7.58
9	14.24
8.5	20.74
8	27.16
7.5	33.54
7	39.87
6.5	46.18
6	52.47

So, $t_{01} = [(10 - 9.5) \div 0.2]^{1/2} = 1.58$ months, and $t_1 = 6 + 1.58 = 7.58$ months.
Proceeding one gets: $e_2 = e_0 - 2 \times \Delta e$, and $e_2 = 10 - 2 \times 0.5 = 9$ mm.
So, $t_{12} = [(10 - 9.0) \div 0.2]^{1/2} = 2.24$ months, and $t_2 = 2 \times 6 + 2.24 = 14.24$ months.
And: $e_3 = e_0 - 3 \times \Delta e$, and $e_3 = 10 - 3 \times 0.5 = 8.5$ mm.
So, $t_{23} = [(10 - 8.5) \div 0.2]^{1/2} = 2.74$ months, and $t_3 = 3 \times 6 + 2.74 = 20.74$ months.

Or, in next Table 1:

Thus, obtaining the instants, or the months from the state of the new sheet metal, in which, foreseeably, the paintings must take place (second column of the previous table). The replacement of the sheet metal should take place at time $t_s = 52.47$ months. Let us now calculate the pertinent costs. However, before proceeding, an effective monthly interest rate given that the annual interest rate i_p is known must be found. The monthly interest rate i_p is equal to [8]:

$i_p = (1 + i)1/f - 1 = (1 + 0.24)1/12 - 1 = 1.81\%$ month

where: i = annual interest rate, and f = frequency of capitalization in a year (12 times in the present case).

So, looking at the uniform costs of painting and replacing the sheet metal, e.g., the uniform cost of a substitution that will take place within 52.47 months will be:

$$A_s = 15{,}000 \times (A/P;\ 1.81;\ 52.47) \cong 445\,€/\text{month}$$

And the uniform cost of painting A_p until then:

$$\begin{aligned} A_P = 10{,}000 \times [&(P/F;\ 1.81;\ 7.58) + (P/F;\ 1.81;\ 14.24) \\ &+ (P/F;\ 1.81;\ 20.74) + \ldots + (P/F;\ 1.81;\ 46.18)] \\ &\times (A/P;\ 1.81;\ 52.47) \cong 1313\,€/\text{month} \end{aligned}$$

So, the total uniform cost A_T will be:

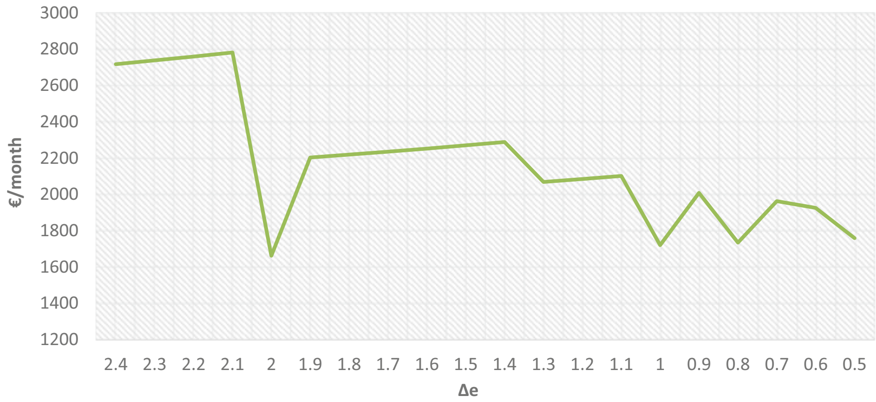


Fig. 5 Monthly cost variation with the magnitude of Δe

$$A_T = A_s + A_p = 445 + 1313 = 1758\,€/\text{month}$$

The total costs of other alternatives of Δe must be calculated. One starts with $\Delta e = 0.5$ mm, which is the smallest value that can be measured, continuing with increments of 0.1 mm, if at least one painting operation will be carried out. After computing the rationale in an EXCEL® spread sheet, the result is depicted in a graph (Fig. 5).

As the graph shows, the minimum monthly cost is achieved for $\Delta e = 2$ mm. In Fig. 5, the graph of the total uniform cost of the twenty alternatives of Δe show a few points of sudden inflection (peaks). These points exist since the number of painting operations per unit of time does not vary continuously over the life span of the sheet metal, as the last painting operation is carried out always when it is noticed that the thickness of the sheet metal at that instant is less than ($e_m + \Delta e$). In fact, searching, for instance, the interval encompassing points $\Delta e = 1.9$, 2, and 2.1 mm, the following reasons are found:

- When $\Delta e = 2.1$ mm, the first and only painting operation would take place at 9.24 months (when the sheet metal had reached the thickness of $10 - 2.1 = 7.9$ mm). The uniform cost of the single painting operation and the cost of replacing the plate at time 10.47 months (when it had reached the thickness of 6 mm) would be, respectively, 1004 €/month and 1777 €/month.
- When $\Delta e = 2.0$ mm, the first and only painting operation would take place at 9.16 months (when the plate had reached the thickness of $10 - 2 = 8$ mm). The uniform cost of the single painting operation and the cost of replacing the plate at the time 16.47 months (when it had reached the thickness of 6 mm) would be, respectively, 600 €/month and 1061 €/month, and
- When $\Delta e = 1.9$ mm, the first painting operation would take place at 9.08 months (when the plate had reached the thickness of $10 - 1.9 = 8.1$ mm) and the second operation at 16.36 months (when the plate had reached the thickness of $10 - 2 \times$

1.9 = 6.2 mm). The uniform cost of these two painting operations and the cost of replacing the plate at the time 16.36 months (when it had reached the thickness of 6 mm) would be respectively 1136 €/month and 1067 €/month.

That is, while the uniform cost of replacement of the sheet metal is progressively reduced by the lengthening of the useful life, the cost of the painting operations is influenced by the variation of the number of paintings per unit of time and therefore oscillates. It is interesting to note that, for the same reason, there are still two other singular points, which are: $\Delta e = 0.8$ and 1 mm.

4 Discussion

Although conceptually simple, since only mathematical concepts of the degrading function, financial conversion, interest rate and optimization were used, the calculations proved to be extensive and somewhat complex, advising the use of automatic calculation, for example in a spreadsheet—today available to any technician or manager. It should be noted that one assumes that the degradation function would be resumed at the point where it had been suspended each time the sheet metal is cleaned and painted. However, this may not always be the real situation. It was not one's intention to consider considerations of this nature, but only to demonstrate the usefulness of combining the descriptive functions of physical phenomena in the field of engineering, with functions of financial analysis, thus completing what one might dubbed, properly, of technical–economic analysis. The search for solutions and forecasting in this context constitutes, in the context of scarce resources, an obviously necessary practice. The case treated in this article allows to highlight how the tools of financial analysis can complete future decisions in engineering practices. Indeed, it does not seem evident that any empirical rule could replace the method one used in its resolution. With one advantage: the method allows generalization and thus the moment of analysis may be any of convenience. That is, with the aid of a short computer program, at any time a measurement of the sheet metal thickness is made, and its value is entered in a specifically developed software, the forecast of how much time is remaining before the next painting operation must take place is obtained.

Acknowledgements The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284 project SSHARE.

References

1. Schuman, C., Brent, A.: Asset life cycle management: towards improving physical asset performance in the process industry. Int. J. Oper. Prod. Manag. **25**(6), 566–579 (2007)
2. Jong, S., Hoefnagels, R., Wetterlund, E., Petterson, K.: Cost optimization of biofuel production—the impact of scale, integration, transport, and supply chain configurations. Appl. Energy **195**, 1055–1070 (2017)
3. Rabin, M.: Incorporating limited rationality into economics. J. Economic Lit. **51**(2), 528–543 (2003)
4. Gomes, F., Michaelides, A.: Optimal life-cycle asset allocation: Understanding the empirical evidence. J. Financ. **60**(2), 869–904 (2005)
5. Shifler, D.: Understanding material interactions in marine environments to promote extended structural life. Corros. Sci. **47**(10), 2335–2352 (2005)
6. Munk, T., Kane, D., Yebra D.: The effects of corrosion and fouling on the performance of ocean-going vessels: a naval architectural perspective. In: Hellio, C., Yebra, D. (eds.) Advances in Marine Antifouling Coatings and Technologies, pp. 148–176 (2009)
7. Holroyd, H., Scamans, G.: Environmental degradation of marine aluminium alloys—past, present, and future. Corrosion **72**(2), 136–143 (2016)
8. Assis, R.: Decision making in maintenance Management of Physical Assets (in Portuguese). Lidel (2014)
9. Canada, J., Sullivan, W., Kulonda, D., White, J.: Capital Investment Analysis for Engineering and Management (3rd ed.). Prentice Hall (2004)
10. Sulivan, W., Wicks, E., Koelling, P.: Engineering Economy (17th ed.). Pearson Education (2018)

Task Allocation Problem Between Human–robot Collaboration Team

Ahmed Abide Tadesse, Kung Jeng Wang, and Chiuhsiang Joe Lin

Abstract The advent of industry 4.0 facilitates the use of collaborative robots near humans. But this proximity brings about some critical challenges. The task allocation problem is one of the top challenges that deserve the serious attention of researchers to attain a successful human–robot collaboration. The main aim of this study is to summarize the existing literature on the topic of the task allocation problem in the human–robot collaboration context. In this paper, we reviewed the task allocation problem from the perspectives of definitions and terminologies, different types of allocation methods, evaluation criteria, implementation procedures, and application phases.

Keywords Industry 4.0 · Human robot collaboration · Task allocation problem

1 Introduction

Researchers and practitioners in developed countries are moving forward in adopting the new paradigm shift named industry 4.0. It is a swift change from the traditional automation of mass production systems to the new era of customized production with the help of autonomous machines and whole inclusive automation. It can be achieved through the Internet of Everything (IoE) as in the ideal cyber-physical production systems [1]. Even though the core idea of industry 4.0 is trying to liberate machines from the intervention of a human, still in some areas specially in assembly operations the role of the human is almost impossible to replace [2]. On the other hand, the current market behavior is trending into mass customization. Which means automation of every production process is out of the window.

A. A. Tadesse (✉) · K. J. Wang · C. J. Lin
National Taiwan University of Science and Technology, Taipei, Taiwan
e-mail: wubied@gmail.com

A. A. Tadesse
Bahir Dar University (BiT), Bahir Dar, Ethiopia

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_45

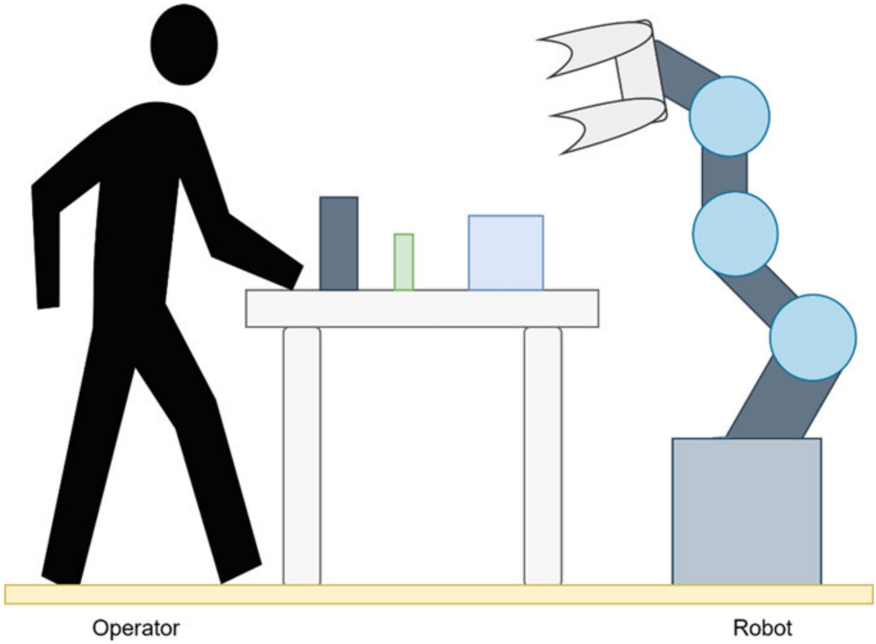


Fig. 1 Human robot collaboration in assembly operation setting

As a result, the immediate solution for such problems would be using manual operation. However, many developed countries are facing an acute shortage of trained assembly workers [3, 4]. Thus, instead of fully-relying on manual assembly or automation, a hybrid production system (see Fig. 1) comprises of robots and humans as a team is the best solution to the challenges faced by modern manufacturers [3, 4].

Collaborative robots have a wide range of applications in wide area of sectors such as: industrial assembly operations, medicine, teaching and learning, elderly and aged care, emotion feedback systems, entertainment, space exploration, and military. Assembly operations are welcoming the idea of using collaborative robots in their production process.

A successful human robot collaboration can be achieved by considering the critical requirements of an HRC team. Safety, communication, task allocation, establishing shared goal, learning and trust are among the crucial factors of a successful human robot collaboration team.

The practical implementation of the HRC team in an industrial environment is full of difficulties, especially in assembly operations. Task allocation is among the top six serious challenges identified by researchers and practitioners [5]. Thus, researchers and practitioners in the field of HRC system design and integration should give due consideration to the task allocation problem (TAP) while designing [6]. Task allocation is the assignment of responsibilities for each member of the collaboration team. A common approach to the problem is decomposition of tasks into a sequence of atomic actions and assign the sub task to the most suitable agent after evaluating with some criteria [7].

In this paper, the investigators discussed task allocation problem from the problem definition, the different types of allocation methods, evaluation criteria, implementation procedures and application phase perspectives in the contemporary literature.

2 A Task Allocation Problem

Task allocation problem is the extension of the traditional task assignment problem in the field of combinatorial optimization. In task assignment the objective is assigning two or more tasks to two or more agents available for assignment having in mind achieving total assignment cost as minimum as possible. Task assignment problem would start first from a given job by analyzing tasks that can be assigned to operator, robot or both. While assigning a multitude of criteria would be considered (see Fig. 2).

2.1 Task Allocation Concept and Terminologies

In the literature, different researchers used twelve terminologies while addressing the task allocation problem (Refer Fig. 3). This paper suggested six categories and contexts that make the terms in different research outputs contextualized. First, researchers used; task *allocation, task assignment, activity allocation, work allocation, and task distribution to refer the general concept of TAP.* The term applied in

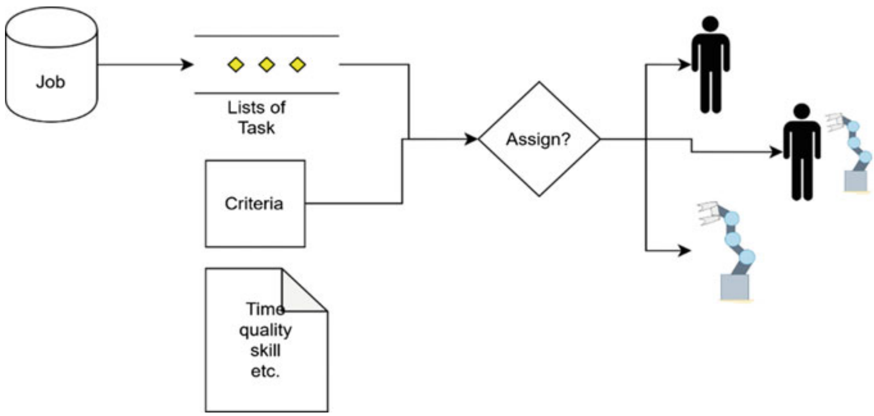


Fig. 2 A generic task assignment procedure

the second category is "mission assignment"; which represents the assignment of a specific goal for an agent. While in the third category, "*role assignment"* is referring to the authority of an agent over the other team member. Whereas; in the fourth category, "task allocation (fixed)" and "dynamic task allocation" signifies the task allocation period and updating conditions. Fifth, "task-allocation", "subtask allocation", and " Task decomposition " are terms used to indicate the hierarchy of a task. Lastly, the phrases "task scheduling" and "online task rescheduling" represents the task sequence and timing considerations.

Fig. 3 Classification of human robot task allocation problem terminologies

Table 1 Task allocation methods

No	Methods	Authors
1	Optimization techniques Multi objective Stochastic integer programming	[2–29]
2	Heuristics Genetic algorithm	[7, 22, 28]
3	Decision making approach Multi criteria decision making Markov decision process	[2, 3, 6, 9, 11, 13, 14, 17, 21, 23, 25, 30, 31]
4	Task analysis approach Task classification	[3, 15, 21, 23, 27, 29]

3 Application Phases and Task Allocation Methods

3.1 Task Allocation Application Phases

Task allocation problem could be studied at different stages of a business life cycle. Whenever an initial investment is considered, TAP can be studied while appraising technical, economic, and market feasibility. If the firm is in the design phase after having a confirmation from the feasibility stand point, then here again TAP should be addressed in relatively deeper sense (see Fig. 4). At this phase TAP should be considered while dealing with system design, layout design, and product design. On the other hand, an existing business may initiate a project to shift from manual operation to HRC. At this phase TAP should be tackled in greater depth with the consideration of the maximum possible factors (see Table 2).

3.2 Task Allocation Methods

In the existing literature, optimization and decision-making approaches are most widely used methods of task allocation. While, other methods such as heuristics, task and system analysis approaches are less utilized (see Table 1).

4 Evaluation Criteria of TAP and Verification Methods

4.1 Evaluation Criteria

In this section, we have presented the criteria used to evaluate and decide a task allocation between human and robot. After surveying the existing literature, we find

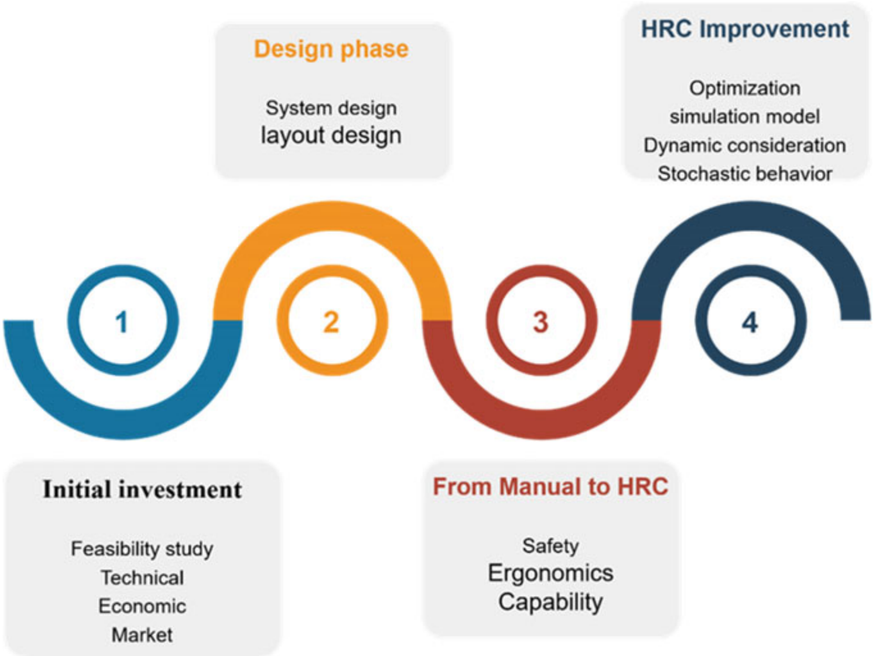


Fig. 4 Task allocation in different phases of business cycle

out 28 criteria employed by researchers while addressing task allocation problem (see Table 2). We analyzed and formulated them in to eight major factors. Time factor, capability factor and ergonomic factor are the most popular and widely used factors among researchers. Safety, production system factor and cost has got moderate attention. Fewer researchers used work place and task factors as a criterion for their evaluation.

4.2 Implementation and Verification Methods

As shown below in Table 3; simulation, experiment, pilot study, case study, proof of concept, numerical example and industrial application are the verification methods widely presented in the literature. Simulation and experimental verification methods are the most widely used techniques for model validation.

Table 2 Task allocation assessment criteria

No	Criteria	Author
1	*Ergonomics Factor* -Agent effort -Personalization -Cognitive load -Trust	[2, 9, 10, 19, 29, 30]
2	*Safety Factor*	[2, 9, 10, 19, 29, 30]
3	*Agent Capability Factor* -Dexterity -Skill -Knowledge -Competences -Quality -Contribution to goal -Work load -Experience -Preference	[2, 9, 10, 19, 29, 30]
4	*Task Factor* -Suitability -Complexity	[3, 15, 21, 27, 29]
5	*Time Factor* -Cycle time -Deadline -Lead time -Completion time -Production time	[26]
6	*Cost Factor*	[2, 9, 10, 19, 29, 30]
7	*Work place Factor* -Adaptability -Personalization	[2, 9, 10, 19, 29, 30]
8	*Production system Factor* -Feasibility -Production Rate -Failure rate -Performance	[2, 9, 10, 19, 29, 30]

Table 3 Model verification methods

No	Methods	Authors
1	Simulation	[5, 13, 14, 15, 17, 18, 29, 30]
2	Experiment	[2, 3, 8, 9, 11, 12, 16, 21, 22, 24]
3	Case/Pilot study	[1, 6, 10, 17, 19, 25]
4	Theory/Proof of concept	[4, 20, 26]
5	Numerical example	[27, 28, 31, 32]
6	Industrial application	[7, 12]

5 Conclusions

Human–robot collaboration is becoming a viable solution for assembly operations. In this paper, we have reviewed the task allocation problem in the human–robot collaboration system. The task allocation problem is about assigning the role and responsibility of an agent. In this paper, we have summarized 14 terminologies used in the literature into six major categories based on the concept, goal, authority, allocation and updating period, task scheduling and sequencing, and task hierarchy. This survey identified the four stages of a business cycle: initial investment, design, system change, and improvement; TAP studied. Researchers used a variety of evaluation criteria while assigning tasks between humans and robots. We have categorized into eight distinct groups.

This study exhibited that simulation and experimental methods are the widely adopted method by researchers. Practical application of HRC in industrial scenario is scarce so future studies could use this opportunity to further investigate TAP.

References

1. Ansari, F., Hold, P., Khobreh, M.: A knowledge-based approach for representing jobholder profile toward optimal human–machine collaboration in cyber physical production systems. CIRP J. Manuf. Sci. Technol. **28**, 87–106 (2020). https://doi.org/10.1016/j.cirpj.2019.11.005
2. Maurtua, I., Ibarguren, A., Kildal, J., Susperregi, L., Sierra, B.: Human–robot collaboration in industrial applications. Int. J. Adv. Rob. Syst. **14**(4), 172988141771601 (2017). https://doi.org/10.1177/1729881417716010
3. Chen, F., Sekiyama, K., Fukuda, T.: A genetic algorithm for subtask allocation within human and robot coordinated assembly. Int. Symp. Micro-Nano Mechatron. Hum. Sci. (MHS) **2012**, 507–511 (2012). https://doi.org/10.1109/MHS.2012.6492504
4. Mateus, J. C., Claeys, D., Limère, V., Cottyn, J., & Aghezzaf, E.-H.: A structured methodology for the design of a human-robot collaborative assembly workplace. Int. J. Adv. Manuf. Technol. **102**(5–8), 2663–2681 (2019). https://doi.org/10.1007/s00170-019-03356-3
5. Ranz, F., Hummel, V., Sihn, W.: Capability-based task allocation in human-robot collaboration. Procedia Manuf. **9**, 182–189 (2017). https://doi.org/10.1016/j.promfg.2017.04.011
6. Malik, A.A., Bilberg, A.: Complexity-based task allocation in human-robot collaborative assembly. Ind. Robot. **46**(4), 471–480 (2019). https://doi.org/10.1108/IR-11-2018-0231
7. Johannsmeier, L., Haddadin, S.: A hierarchical human-robot interaction-planning framework for task allocation in collaborative industrial assembly processes. IEEE Robot. Autom. Lett. **2**(1), 41–48 (2017). https://doi.org/10.1109/LRA.2016.2535907
8. Alami, R., Clodic, A., Montreuil, V., Sisbot, E.A., Chatila, R.: Task planning for human-robot interaction. ACM Int. Conf. Proc. Ser. **121**, 81–85 (2005). https://doi.org/10.1145/1107548.1107574
9. Banerjee, A.G., Barnes, A., Kaipa, K.N., Liu, J., Shriyam, S., Shah, N., Gupta, S.K.: An ontology to enable optimized task partitioning in human-robot collaboration for warehouse kitting operations. In Popa, D., Wijesundara, M.B.J., Blowers, M. (eds.) Next-generation robotics II; and machine intelligence and bio-inspired computation: theory and applications IX, vol. 9494(18), p. 94940H. SPIE. https://doi.org/10.1117/12.2181346
10. Blankemeyer, S., Wiemann, R., Vett, U.K.: Efficient Use of Human-robot collaboration in packaging through systematic task assignment. https://doi.org …, 297–306. https://www.repo.uni-hannover.de/handle/123456789/9727

11. Bogner, K., Pferschy, U., Unterberger, R., Zeiner, H.: Optimised scheduling in human–robot collaboration–a use case in the assembly of printed circuit boards. Int. J. Prod. Res. **56**(16), 5522–5540 (2018). https://doi.org/10.1080/00207543.2018.1470695
12. Bruno, G., & Antonelli, D.: Dynamic task classification and assignment for the management of human-robot collaborative teams in workcells. Int. J. Adv. Manuf. Technol. **98**(9-12), 2415–2427 (2018). https://doi.org/10.1007/s00170-018-2400-4
13. Bänziger, T., Kunz, A., Wegener, K.: Optimizing human–robot task allocation using a simulation tool based on standardized work descriptions. J. Intell. Manuf. **31**(7), 1635–1648 (2020). https://doi.org/10.1007/s10845-018-1411-1
14. Chen, F., Sekiyama, K., Huang, J., Sun, B., Sasaki, H., Fukuda, T.: An assembly strategy scheduling method for human and robot coordinated cell manufacturing. Int. J. Intell. Comput. Cybern. **4**(4), 487–510 (2011a). https://doi.org/10.1108/17563781111186761
15. Chen, F., Sekiyama, K., Sasaki, H., Huang, J., Sun, B., Fukuda, T.: Assembly strategy modeling and selection for human and robot coordinated cell assembly. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4670–4675 (2011b). https://doi.org/10.1109/IROS.2011.6048306
16. Chen, F., Sekiyama, K., Cannella, F., & Fukuda, T.: Optimal Subtask Allocation for Human and Robot Collaboration Within Hybrid Assembly System. IEEE Trans. Autom. Sci. Eng. **11**(4), 1065–1075 (2014). https://doi.org/10.1109/tase.2013.2274099
17. Costa Mateus, J.E., Aghezzaf, E.H., Claeys, D., Limère, V., Cottyn, J.: Method for transition from manual assembly to Human-Robot collaborative assembly. IFAC-PapersOnLine, **51**(11), 405–410 (2018). https://doi.org/10.1016/j.ifacol.2018.08.328
18. Ding, H., Schipper, M., Matthias, B.: Optimized task distribution for industrial assembly in mixed human-robot environments—Case study on IO module assembly. In: 2014 IEEE International Conference on Automation Science and Engineering (CASE), 19–24 Jan 2014 (2014). https://doi.org/10.1109/CoASE.2014.6899298
19. Djezairi, S., Akli, I., Zamoum, R.B., Bouzouia, B.: Mission allocation and execution for human and robot agents in industrial environment. In: RO-MAN 2018—27th IEEE International Symposium on Robot and Human Interactive Communication, pp. 796–801 (2018). https://doi.org/10.1109/ROMAN.2018.8525726
20. Gualtieri, L., Rauch, E., Vidoni, R., & Matt, D. T.: An evaluation methodology for the conversion of manual assembly systems into human-robot collaborative workcells. Procedia Manuf. **38**, 358–366 (2019). https://doi.org/10.1016/j.promfg.2020.01.046
21. Gualtieri, L., et al., Implementation of a laboratory case study for intuitive collaboration between man and machine in SME assembly, in Industry 4.0 for SMEs. Palgrave Macmillan, Cham. p. 335-382 (2020). https://doi.org/10.1007/978-3-030-25425-4_12
22. Lamon, E., de Franco, A., Peternel, L., Ajoudani, A.: A capability-aware role allocation approach to industrial assembly tasks. IEEE Rob. Autom. Lett. **4**(4), 3378–3385 (2019). https://doi.org/10.1109/LRA.2019.2926963
23. Pellegrinelli, S., Orlandini, A., Pedrocchi, N., Umbrico, A., Tolio, T.: Motion planning and scheduling for human and industrial-robot collaboration. CIRP Ann. Manuf. Technol. **66**(1), 1–4 (2017). https://doi.org/10.1016/j.cirp.2017.04.095
24. Rahman, S. M., Sadrfaridpour, B., & Wang, Y.: Trust-based optimal subtask allocation and model predictive control for human-robot collaborative assembly in manufacturing. *Dynamic Systems and Control Conference. American Society of Mechanical Engineers,* (2015). https://doi.org/10.1115/DSCC2015-9850
25. Riedelbauch, D., Henrich, D.: Exploiting a human-aware world model for dynamic task allocation in flexible human-robot teams. In: Proceedings—IEEE International Conference on Robotics and Automation, 6511–6517 May (2019). https://doi.org/10.1109/ICRA.2019.8794288
26. Roncone, A., Mangin, O., Scassellati, B.: Transparent role assignment and task allocation in human robot collaboration. In: Proceedings—IEEE international conference on robotics and automation, pp. 1014–1021 (2017). https://doi.org/10.1109/ICRA.2017.7989122

27. Stohr, M., Schneider, M., Henkel, C.: Adaptive work instructions for people with disabilities in the context of human robot collaboration. In: Proceedings—IEEE 16th International Conference on Industrial Informatics, INDIN 2018, pp. 301–308 (2018). https://doi.org/10.1109/INDIN.2018.8472070
28. Takata, S., Hirano, T.: Human and robot allocation method for hybrid assembly systems. CIRP Ann. **60**(1), 9–12 (2011). https://doi.org/10.1016/j.cirp.2011.03.128
29. Tsarouchi, P., Michalos, G., Makris, S., Athanasatos, T., Dimoulas, K., Chryssolouris, G.: On a human–robot workplace design and task allocation system. Int. J. Comput. Integr. Manuf. **30**(12), 1272–1279 (2017). https://doi.org/10.1080/0951192X.2017.1307524
30. Tsarouchi, P., Spiliotopoulos, J., Michalos, G., Koukas, S., Athanasatos, A., Makris, S., Chryssolouris, G.: A decision making framework for human robot collaborative workplace generation. Procedia CIRP **44**, 228–232 (2016). https://doi.org/10.1016/j.procir.2016.02.103
31. Weßkamp, V., Seckelmann, T., Barthelmey, A., Kaiser, M., Lemmerz, K., Glogowski, P., Kuhlenkötter, B., Deuse, J.: Development of a sociotechnical planning system for human-robot interaction in assembly systems focusing on small and medium-sized enterprises. Procedia CIRP **81**, 1284–1289 (2019). https://doi.org/10.1016/j.procir.2019.04.014
32. Wilcox, R.J.: Flexible schedule optimization for human-robot collaboration. Massachusetts Institute of Technology (2013). https://dspace.mit.edu/handle/1721.1/81713

An Exact Algorithm for the Monitoring Problem by Using Drones

Gwang Kim, Jongmin Lee, and Ilkyeong Moon

Abstract Drones are recently used for monitoring incidents in smart cities such as crime, transportation, disaster, etc. In this study, we present a monitoring problem by using drones. The monitoring problem is represented as a submodular maximization problem with a partition matroid constraint. The problem is known as NP-hard in general. If incidents that come up in a city lead to severe consequences, more effective solutions to monitor the incidents are needed to prevent as much damage as possible from the incidents. We develop an exact algorithm to solve the monitoring problem. We present two types of valid inequalities that are used to construct the exact algorithm. We also prove that the two types of the valid inequalities are valid in the monitoring problem. By using (near)-optimal solutions of drone operations from the exact algorithm, we can minimize losses from the occurrences of incidents. Numerical tests are also presented to show the performance of the algorithm.

Keywords Drone · Monitoring · Submodular maximization · Valid inequality · Exact algorithm

G. Kim (✉)
Division of Business Administration, Chosun University, Gwangju, Korea
e-mail: gwangkim91@chosun.ac.kr

J. Lee · I. Moon
Department of Industrial Engineering, Seoul National University, Seoul, Korea
e-mail: jmlee0@snu.ac.kr

I. Moon
e-mail: ikmoon@snu.ac.kr

I. Moon
Institute for Industrial Systems Innovation, Seoul National University, Seoul, Korea

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_46

1 Introduction

Drone, known as Unmanned Aerial Vehicle (UAV), is a type of aircraft that has no human pilot on board. Drones are used in diverse areas such as the military, agriculture, and so on [1, 2]. In addition to the areas, city administrators have been considering drones for monitoring incidents that come up within the city such as pollution, transportation, and disaster [3, 4]. As the size of a city grows and the concept of a smart city appears, administrators have been starting to find the appropriate ways to monitor the city. Drones are becoming an implemental way for monitoring big cities. Operations by using drones would be efficient because drones have the property of moving around in the air contrary to stationary sensors [5]. It is possible to monitor the city with a relatively small number of drones. As the size of the city to monitor is bigger and the number of drones used increases, it becomes more challenging to properly operate drones to achieve effective monitoring.

In this study, we present a monitoring problem by using drones. Given a two-dimensional space, we consider a set of drones to monitor and a set of targets (locations) to be monitored. During a period, each drone moves to a point from the previous point. The distances between the two points are limited because the maximum distance for each drone in one period is set depending on drone specification, weather, city environment, etc. After each drone moves, it monitors a subset of targets that are positioned within its own bounded monitoring radius from the drone. The problem in this study follows Elfes sensing model, which is one of the probabilistic sensing models, to represent utility obtained from monitoring [6]. Each target gets its own utility depending on the monitoring situation from the drones. As the distance between a drone and a target gets closer, the utility gets increased. The objective of the monitoring problem by using drones is to maximize the sum of utilities obtained from all targets in a period. Each drone has possible points in each period depending on the previous point of the drone. The solutions to the problem become a set of located points of all drones that are moved from the previous point.

The objective function of the problem is nondecreasing and submodular. A submodular function has a property of the marginal gain diminishing, which can correspond to the discrete counterpart to concave functions [7]. There has been literature on the monitoring or sensing problems considering submodularity [8, 9]. The constraint of the problem is to choose each point of all drones from all possible points, which corresponds to a partition matroid. Therefore, the monitoring problem by using drones is represented as a submodular maximization problem with a partition matroid constraint. The problem is generally known as NP-hard [10].

There have been studies on monitoring problem that is related to a submodular maximization problem. Krause et al. [11] and Shamaiah et al. [12] proposed a sensor placement (selection) problem to monitor large networks. The objective functions of the research follow submodularity. In Rezazadeh and Kia [13], a problem of persistent monitoring for event detection is proposed and a sequential greedy algorithm is presented to deal with the submodularity of the objective function. Most of the literature tends to present algorithms to deal with monitoring problems that consider submodularity. The authors obtained approximation solutions of greedy-type algorithms and proved the performance guarantees by using the property of submodularity [14–16]. However, if incidents that come up in a city lead to severe consequences, more effective solutions to monitor the incidents are needed to prevent as much damage as possible from the incidents. In this study, an exact algorithm to obtain (near)-optimal solutions to the monitoring problem is presented. We design the algorithm by using two types of valid inequalities and show the performances of the algorithm through numerical experiments.

This paper is organized as follows. Section 2 explains the problem statements and describes the monitoring problem by using drones mathematically. In Sect. 3, we present two types of valid inequalities and an exact algorithm by using the valid inequalities. Section 4 provides numerical experiments to validate the algorithm. We conclude the paper in Sect. 5.

2 Problem Description

In this section, we define a monitoring problem to maximize the sum of the utilities obtained from all targets in a period. We consider a set of drones $D = \{1, \ldots, N\}$ and a set of targets $T = \{1, \ldots, M\}$. Each drone moves to a point in a given space and monitors a subset of targets that are positioned within its own bounded monitoring radius from the drone. It is assumed that all targets are stationary and target t is located at O_τ. . We decide the next points of drones based on the previous points of the drones. Let the (next) points that drones are located be $L_D = \{l_1, \ldots, l_N\}$. It is also assumed that each drone has its own maximum distance that a drone can move during a unit period and drone d has a monitoring range of r_d. Based on the Elfes sensing model [6], the probability that drone d detects an incident at target t is defined as follows:

$$p(l_d, t) = \begin{cases} \exp(-\lambda_d \|l_d - O_t\|), & \text{if } \|l_d - O_t\| \le r_d \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where l_d is the point of drone d and λ_d is a detection decay factor of drone d. There might be more than two drones that monitor the same target. This means that we calculate the joint utility of target t as follows:

$$U_t(L_D) = 1 - \prod_{l_d \in L_D} (1 - p(l_d, t)) \tag{2}$$

where we assume that detection probabilities of drones are independent. The objective of the problem is to maximize the sum of utilities obtained from all targets in a period. Therefore, we design the objective function of the monitoring problem by using drones as follows:

$$\max_{L_D} U(L_D) = \sum_t U_t(L_D) \tag{3}$$

Each drone is allowed to move to only one point among the possible points, which can be a constraint of the problem. The constraint is represented as a condition for partition matroid [10].

The objective function of the problem is a non-decreasing and submodular set function. There is a set function $f : 2^X \to \mathbb{R}.$. Given any two sets (A and B) such as $A \subset B \subseteq X$ and an element $w \notin B$, $U(\cdot)$ is submodular if $f(B \cup \{w\}) - f(B) \leq f(A \cup \{w\}) - f(A).$. This paper follows a similar manner of proof procedure that objective function (3) is non-decreasing and submodular in Theorem 1 of Lee et al. [16].

3 Exact Algorithm

As shown in Sect. 2, the monitoring problem by using drones corresponds to a submodular maximization problem with a partition matroid constraint. Because the problem is generally known as NP-hard, it might be difficult to obtain optimal solutions within a reasonable time. This paper presents an exact algorithm to obtain optimal solutions to the problem. To solve the problem efficiently, we present two types of valid inequalities and use the inequalities to construct an exact algorithm. Before dealing with the inequalities, we focus on the partition matroid constraint of the problem. Instead of the set function, we represent the constraint as a form of integer programming to present the valid inequalities. We design the form to match the set function (L_D) with an IP expression. If drone d is allowed to move to n_d points from the previous point in a period, we can define n_d decision variables ($x_{d1}, x_{d2}, \ldots, x_{dn_d}$). If the drone moves to kth point among n_d points, we set $x_{dk} = 1$; otherwise 0. Drone d has to satisfy $\sum_{j=1}^{n_d} x_{dj} = 1$ which is equal to a partition matroid constraint. We present a mathematical formulation (Formulation 1) using $\sum_{j=1}^{n_d} x_{dj} = 1$ for all drones ($d \in D$). The formulation below will be used in the first iteration of the exact algorithm.

[Formulation 1]

$$\max \quad ub$$

$$\text{s.t.} \quad \sum_{j=1}^{n_d} x_{dj} = 1, \forall d \tag{4}$$

$$x_{dj} \in \{0, 1\}, \forall d, j \text{ and ub} \in \mathbb{R} \tag{5}$$

where *ub* is the objective value of Formulation 1 and plays a role in an upper bound of the monitoring problem. In the first iteration of the exact algorithm, the value of

ub becomes $+\infty$. As valid inequalities related to ub are added as constraints, the upper bound gets decreased.

Using the decision variables, we present two types of valid inequalities based on a feasible solution $\overline{X}$ that is an IP expression of set L_D. $\overline{X} = \left\{\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_N\right\}$ and L_d corresponds to $\overline{X}_d$ for all drones ($d \in D$). Let X_d be a set of feasible solutions of drone d. We use x_{dk} such that $x_{dk} = 1$ for a given $\overline{X}$ in the inequalities below.

$$ub \leq U(\overline{X}) + \sum_{d=1}^{N} \left[U\left(\overline{X} \cup \{w_d\}\right) - U(\overline{X})\right] * (1 - x_{dk}) \tag{6}$$

$$ub \leq U(\overline{X}) + \sum_{d=1}^{N} \left[U(\{v_d\}) - \left\{U(\overline{X}) - U\left(\overline{X} \backslash \overline{X}_d\right)\right\}\right] * (1 - x_{dk}) \tag{7}$$

where $w_d = \text{argmax}_w U(\overline{X} \cup \{w\}) - U(\overline{X})$ and $v_d = \text{argmax}_v U(\{v\})$ such that $w, v \in X_d, \forall d \in D$. We prove that the two types of valid inequalities are valid in the monitoring problem.

3.1 Proofs of Valid Inequalities

We prove that Inequality (6) by using a feasible solution $\overline{X}$ is valid in the monitoring problem. First of all, we know that $U(\cdot)$ is submodular. In Nemahuser et al. [10], if $U(\cdot)$ is submodular, then $U(B) \leq U(A) + \sum_{w \in B \backslash A}[U(A \cup \{w\}) - U(A)]$ is satisfied. In Formulation 1, there exists an arbitrary feasible solution (X, ub) such that $ub \leq U(X)$.

$$\begin{aligned} ub &\leq U(X) \\ ub &\leq U(\overline{X}) + \sum_{w \in X \backslash \overline{X}} [U(\overline{X} \cup \{w\}) - U(\overline{X})] \end{aligned} \tag{8}$$

$$ub = U(\overline{X}) + \sum_{d=1}^{N} \sum_{w \in X_d \backslash \overline{X}_d} [U(\overline{X} \cup \{w\}) - U(\overline{X})]$$

$$ub = U(\overline{X}) + \sum_{d=1}^{N} \sum_{w \in X_d \backslash \overline{X}_d} [U(\overline{X} \cup \{w\}) - U(\overline{X})] * (1 - x_{dk}) \tag{9}$$

Inequality (8) is from $U(B) \leq U(A) + \sum_{w \in B \backslash A}[U(A \cup \{w\}) - U(A)]$. Inequality (9) follows from the definition of w_d. It means that $ub \leq U(\overline{X}) + \sum_{d=1}^{N} \left[U\left(\overline{X} \cup \{w_d\}\right) - U(\overline{X})\right] * (1 - x_{dk}))$ is valid in the monitoring problem.

We also prove that Inequality (4) by using a feasible solution $\overline{X}$ is valid in the monitoring problem. In Nemahuser et al. [10], if $U(\cdot)$ is submodular, then $U(B) \le U(A) + \sum_{w \in B \setminus A}[U((A \cap B) \cup \{v\}) - U(A \cap B)] - \sum_{w \in A \setminus B}[U(A) - U(A \setminus \{v\})]$ is satisfied. As with the proof of Inequality (7), there exists an arbitrary feasible solution (X, ub) such that $ub \le U(X)$.

$$\begin{aligned} ub &\le U(X) \\ ub &\le U(\overline{X}) + \sum_{v \in X \setminus \overline{X}} [U((\overline{X} \cap X) \cup \{v\}) \\ &\quad - U((\overline{X} \cap X))] - \sum_{v \in \overline{X} \setminus X} [U(\overline{X}) - U(\overline{X} \setminus \{v\})] \end{aligned} \tag{10}$$

$$ub \le U(\overline{X}) + \sum_{v \in X \setminus \overline{X}} U(\{v\}) - \sum_{v \in \overline{X} \setminus X} [U(\overline{X}) - U(\overline{X} \setminus \{v\})] \tag{11}$$

$$\begin{aligned} ub &= U(\overline{X}) + \sum\nolimits_{d=1}^{N} \sum_{v \in X \setminus \overline{X}_d} U(\{v\}) \\ &\quad - \left[\sum_{v \in \overline{X}} [U(\overline{X}) - U((\overline{X}) \setminus \{v\}))] - \sum_{v \in \overline{X} \cap X} [U(\overline{X}) - U(\overline{X} \setminus \{v\})] \right] \\ ub &= U(\overline{X}) + \sum_{d=1}^{N} \sum_{v \in X_d \setminus \overline{X}_d} U(\{v\}) * (1 - x_{dk}) \\ &\quad - \left[\sum\nolimits_{d=1}^{N} \sum\nolimits_{v \in \overline{X}_d} [U(\overline{X}) - U(\overline{X} \setminus \{v\}) \right] \\ &\quad - \sum\nolimits_{d=1}^{N} \sum_{v \in \overline{X}_d \cap \overline{X}_d} \left[U\left(\overline{X}\right) U(\overline{X} \setminus \{v\})] * (x_{dk}) \right] \\ ub &\le U(\overline{X}) + \sum_{d=1}^{N} \sum_{v \in X_d \setminus \overline{X}_d} U(\{v\}) * (1 - x_{dk}) \\ &\quad - \left[\sum\nolimits_{d=1}^{N} \sum\nolimits_{v \in \overline{X}_d} [U(\overline{X}) - U(\overline{X} \setminus \{v\}) \right] \\ &\quad - \sum\nolimits_{d=1}^{N} \sum_{v \in \overline{X}_d} \left[U\left(\overline{X}\right) U(\overline{X} \setminus \{v\})] * (x_{dk}) \right] \end{aligned} \tag{12}$$

$$\begin{aligned} ub &\le U(\overline{X}) + \sum_{d=1}^{N} [U(\{v_d\}) * (1 - x_{dk}) \\ &\quad - \left[\sum\nolimits_{d=1}^{N} \sum\nolimits_{v \in \overline{X}_d} [U(\overline{X}) - U(\overline{X} \setminus \{v\}) \right] \\ &\quad - \sum\nolimits_{d=1}^{N} \sum_{v \in \overline{X}_d} \left[U\left(\overline{X}\right) U(\overline{X} \setminus \{v\})] * (x_{dk}) \right] \end{aligned}$$

$$ub = U(\overline{X}) + \sum_{d=1}^{N} \left[U(\{v_d\}) * (1 - x_{dk}) - \sum_{d=1}^{N} \left[U(\overline{X}) - U\left(\overline{X} \backslash \overline{X}_d\right) \right] * (1 - x_{dk}) \right.$$

$$ub = U(\overline{X}) + \sum_{d=1}^{N} \left[U(\{v_d\}) - \left\{ U(\overline{X}) - U\left(\overline{X} \backslash \overline{X}_d\right) \right\} \right] * (1 - x_{dk}) \tag{13}$$

Inequality (10) is from $U(B) \leq U(A) + \sum_{v \in B \backslash A} [U((A \cap B) \cup \{v\}) - U(A \cap B)] - \sum_{v \in A \backslash B} [U(A) - U(A \backslash \{v\})]$. Inequality (11) follows from the submodularity of U(.). Because we know $\left(\overline{X}_d \cap X_d\right) \subseteq \overline{X}_d$ and $U(\cdot)$ is submodular Inequalities (12) is satisfied. Inequality (13) follows from the definition of v_d. It means that $ub \leq U(\overline{X}) + \sum_{d=1}^{N} \left[U(\{v_d\}) - \left\{ U(\overline{X}) - U\left(\overline{X} \backslash \overline{X}_d\right) \right\} \right] * (1 - x_{dk})$ is valid in the monitoring problem. When the two inequalities are added in Formulation 1 as constraints, they play a role in decreasing the upper bound of the monitoring problem. In Sect. 3.2, an exact algorithm is presented by using Formulation 1 and the two types of valid inequalities.

3.2 Exact Algorithm by Using Valid Ineqalities

In this subsection, we develop an exact algorithm to solve the monitoring problem by using the two types of valid inequalities. Contrary to the greedy-type algorithms, the exact algorithm can obtain (near)-optimal solutions to the monitoring problem. It is more useful when incidents that come up in a city lead to severe consequences. The exact algorithm is described as follows.

1. **Initialize**. Let $UB \leftarrow +\infty$ and $LB \leftarrow -\infty$.
2. **Termination check**. If$(UB - LB)/UB \leq \epsilon$, then terminate the algorithm. Otherwise, go to Step [3].
3. Solve Formulation 1. Let $(\overline{X}, ub)$ be the optimal solution of Formulation 1. Then, $UB \leftarrow ub$.
4. If $LB \leq U(\overline{X})$, then $LB \leftarrow U(\overline{X})$.
5. Using $\overline{X}$,, add Inequalities (3) and (4) to Formulation 1 as constraints. Return to Step [2].

Let ϵ be the gap between the upper bound (*UB*) and lower bound (*LB*) of the monitoring problem. In the first iteration of Step [3], when Formulation 1 is solved, we can get an arbitrary solution that is satisfied with feasibility. As the valid inequalities related to *ub* and $\overline{X}$ are added as constraints, the upper bound gets decreased. If the condition of Step [2] is satisfied, $\overline{X}$, which is the current solution of the monitoring problem, becomes the (near)-optimal solution.

4 Numerical Experiments

In this section, numerical tests were conducted on the monitoring problem by using drones. All experiments were performed on a CPLEX version. 20.1. We selected the appropriate number of drones and the number of targets in consideration of the realistic situation. The drones and the targets were randomly generated from points in a two-dimensional space.

First of all, we conducted numerical tests to analyze the process to obtain the optimal solution. When executing the exact algorithm, it reduces the optimality gap between the lower bound and the upper bound by adding two types of valid inequalities for each iteration. To validate this process, we compared the computation times and the number of constraints added where '1- optimality gap (%)' is 60%, 75%, 90%, and 99%, respectively. For the data set, we set $|D| = 10,\ |T| = 20$ and $|D| = 20,\ |T| = 40$ to reflect the real situation in this experiment. We executed 50 runs for each data set and averaged the results. The time limit was set to 3600 s. Figure 1 represents two plots for comparing the number of constraints added when '1-optimality gap (%)' is 60%, 75%, 90%, and 99%, respectively. As a result of the experiment, the solution improved smoothly as the constraints were added until '1-optimality gap (%)' reached 90%. However, we confirmed that considerable computation time and the number of constraints added were required to improve '1-optimality gap (%)' from 90 to 99%, especially in the case of $|D| = 20,\ |T| = 40$. Figure 2 represents two plots for comparing the computation times when '1-optimality gap (%)' is 60%, 75%, 90%, and 99%, respectively. We confirmed a similar tendency as in Fig. 1. When '1-optimality gap (%)' is 99%, it can be seen that the difference in computation time is much larger. In addition, in the all cases of $|D| = 10,\ |T| = 20$, we obtained the optimal solutions less than 10 s.

For the large data set, we set $|D| = 30,\ |T| = 60$ to assume a complex situation in this experiment. Also, to analyze the effect of the initial solution, we compared two methods (Random & Greedy). In Random, we start the exact algorithm by using an arbitrary solution that is satisfied with feasibility. In Greedy, we set an initial

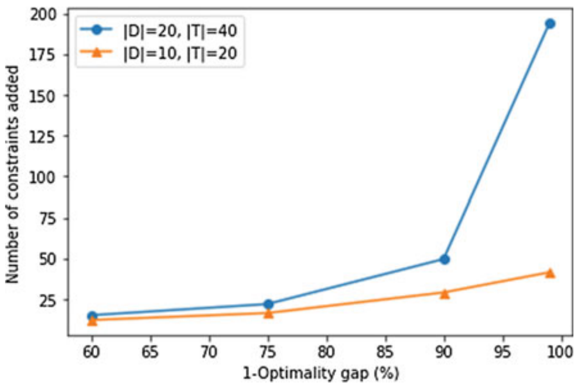


Fig. 1 Comparison of the number of constraints added when '1-optimality gap (%)' is 60%, 75%, 90%, and 99%, respectively

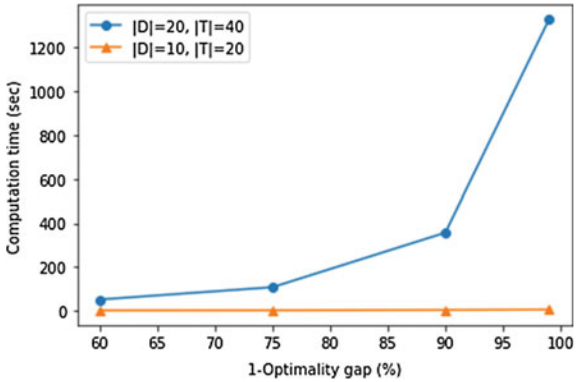


Fig. 2 Comparison of the computation times when '1-optimality gap (%)' is 60%, 75%, 90%, and 99%, respectively

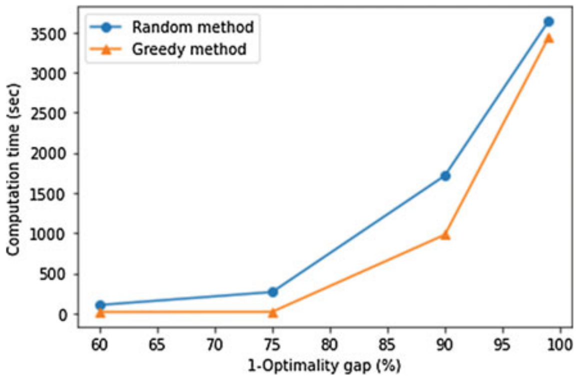


Fig. 3 Comparison of the computation times when '1-optimality gap (%)' is 60%, 75%, 90%, and 99%, respectively

solution based on the greedy algorithm. We executed 10 runs for each method and averaged the results. The time limit was set to 3600 s. Figure 3 represents plots of the two methods for comparing the computation times when '1-optimality gap (%)' is 60%, 75%, 90%, and 99%, respectively. We observed that the computation times to improve the solution increase exponentially. In some cases, '1-optimality gap (%)' did not reach 99% due to the time limit. Comparing the performances of the final solution, on average, was 91.90% in Random and 93.32% (Greedy). Depending on the way to set initial solutions, the performances of the exact algorithm would be different.

5 Conclusions

Drones can be an efficient way to monitor the city because they have the property of moving around in the air. In this study, we presented a monitoring problem by using drones. Because the size of the city is bigger and the consequences of the incidents that come up within the city are more unexpected, it is necessary to propose an effective operation of monitoring by using drones to prevent as much damage as possible. We designed an exact algorithm to solve the monitoring problem. The exact algorithm can obtain (near)-optimal solutions by using two types of valid inequalities. We also proved the validity of the inequalities as constraints. We showed that the performances of the exact algorithm (the computation times and the number of constraints added) through numerical experiments.

For future research, other algorithms to cover the monitoring problem should be developed in terms of effective and efficient performances. Also, there might be uncertain situations such as failure in monitoring targets in the problem. In this case, we can consider the monitoring problem as an application of stochastic programming. The solutions from stochastic programming will become more risk-averse strategies of drone movements.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning [Grant No. NRF2019R1A2C2084616]

References

1. Tokekar, P., Vander Hook, J., Mulla, D., Isler, V.: Sensor planning for a symbiotic UAV and UGV system for precision agriculture. IEEE Trans. Rob. **32**(6), 1498–1511 (2016). https://doi.org/10.1109/TRO.2016.2603528
2. Azzabi, T., Farhat, H., Sahli, N.: A survey on wireless sensor networks security issues and military specificities. In: 2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET), pp. 66–72. IEEE (2017). https://doi.org/10.1109/ASET.2017.7983668
3. Jamil, M.S., Jamil, M.A., Mazhar, A., Ikram, A., Ahmed, A., Munawar, U.: Smart environment monitoring system by employing wireless sensor networks on vehicles for pollution free smart cities. Procedia Eng. **107**, 480–484 (2015). https://doi.org/10.1016/j.proeng.2015.06.106
4. Zhu, F., Li, Z., Chen, S., Xiong, G.: Parallel transportation management and control system and its applications in building smart cities. IEEE Trans. Intell. Transp. Syst. **17**(6), 1576–1585 (2016). https://doi.org/10.1109/TITS.2015.2506156
5. Liu, B., Brass, P., Dousse, O., Nain, P., Towsley, D.: Mobility improves coverage of sensor networks. In: Proceedings of the 6th ACM International Symposium on Mobile ad hoc Networking and Computing, pp. 300–308 (2005). https://doi.org/10.1145/1062689.1062728
6. Elfes, A.: Using occupancy grids for mobile robot perception and navigation. Computer **22**(6), 46–57 (1989). https://doi.org/10.1109/2.30720
7. Calinescu, G., Chekuri, C., Pal, M., Vondrák, J.: Maximizing a monotone submodular function subject to a matroid constraint. SIAM J. Comput. **40**(6), 1740–1766 (2011). https://doi.org/10.1137/080733991
8. Jawaid, S.T., Smith, S.L.: Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems. Automatica **61**, 282–288 (2015). https://doi.org/10.1016/j.automatica.2015.08.022
9. Krause, A., Guestrin, C.: Submodularity and its applications in optimized information gathering. ACM Trans. Intell. Syst. Technol. (TIST) **2**(4), 1–20 (2011). https://doi.org/10.1145/1989734.1989736

10. Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions—I. Math. Program. **14**(1), 265–294 (1978). https://doi.org/10.1007/BF01588971
11. Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., Faloutsos, C.: Efficient sensor placement optimization for securing large water distribution networks. J. Water Res. Planning Manag. **134**(6), 516-526 (2008). 10.1061
12. Shamaiah, M., Banerjee, S., Vikalo, H.: Greedy sensor selection: Leveraging submodularity. In: 49th IEEE Conference on Decision and Control (CDC), pp. 2572–2577, IEEE (2010). https://doi.org/10.1109/CDC.2010.5717225
13. Rezazadeh, N., Kia, S.S.: A sub-modular receding horizon solution for mobile multi-agent persistent monitoring. Automatica **127**, 109460 (2021). https://doi.org/10.1016/j.automatica.2020.109460
14. Qu, G., Brown, D., Li, N.: Distributed greedy algorithm for multi-agent task assignment problem with submodular utility functions. Automatica **105**, 206–215 (2019). https://doi.org/10.1016/j.automatica.2019.03.007
15. Sun, X., Cassandras, C.G., Meng, X.: Exploiting submodularity to quantify near-optimality in multi-agent coverage problems. Automatica **100**, 349–359 (2019). https://doi.org/10.1016/j.automatica.2018.11.020
16. Lee, J., Kim, G., Moon, I.: A mobile multi-agent sensing problem with submodular functions under a partition matroid. Comput. Oper. Res. **132**, 105265 (2021). https://doi.org/10.1016/j.cor.2021.105265

A Two-Level Induced OWA Procedure for Ranking DMUs Under a DEA Cross-Efficiency Framework

Amar Oukil

Abstract Cross-efficiency (CE) evaluation is an extension of data envelopment analysis (DEA) used to fully rank decision-making units (DMUs). The ranking process is carried out on the matrix of CE scores, where an ultimate efficiency score is computed for each DMU through an adequate amalgamation process. In this paper, we propose a ranking procedure that computes the ultimate scores over two aggregation levels, which involve ordered weighted averaging (OWA) induced by different preference settings. First, the preference voting system embedded under the CE matrix is employed to derive aggregate votes as inner quantifiers of the consensual importance of the DMUs. Next, the preference order induced from the aggregate votes is adopted as a ground for rearranging the rows of the original CE matrix prior to the ultimate calculation of the efficiency scores. To substantiate the impact of subjectivity on the structure of the ranking patterns and, hence, assess the robustness of the proposed procedure, the induced OWA is implemented with different optimism level values. The proposed methodology is applied for selecting the best material handling equipments (MHE) within a sample of 25 real-life MHEs, collected from catalogues of MHE manufacturers and vendors.

Keywords Induced ordered weighted averaging · Data envelopment analysis · Preference voting · Aggregation

1 Introduction

The cross-efficiency (CE) [1] data envelopment analysis (DEA) continuous its upsurge as one of the most potent tools for ranking homogeneous decision making units (DMUs) that utilize multiple inputs (resources) to create multiple outputs (products/ services) [2–5]. The application spectrum of CE is broad and includes premium

A. Oukil (✉)
Department of Operations Management & Business Statistics, Sultan Qaboos University, Muscat, Oman
e-mail: aoukil@squ.edu.om

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_47

allocation [6], supply chain management [7], ranking players [8, 9], portfolio selection [10], supplier selection [11], stock selection [12], ranking dispatching rules [13], and selection of material handling equipment [14], to mention only a few recent areas.

Under a CE setting, each DMU uses its weight profile to evaluate its peers and the outcome is typically a vector of peer and self-efficiency scores. The amalgamation of the latter scores produces the ultimate scores needed to rank the DMUs. Examples of amalgamation mechanisms involve nucleolus solutions [15], Shapley values [16], Shannon entropy [17], evidential reasoning [18], Ordered Weighted Averaging (OWA) operators [19], and induced OWA (IOWA) operator [20]. Contrary to OWA, the IOWA enables the scores to be ordered with criteria suitably defined by an order-inducing variable rather than magnitudes. As a result, the ranking patterns that are produced are more consistent and widely accepted.

In our paper, we develop an aggregation procedure that extracts the order-inducing variables from the preference voting system embedded under the CE matrix [21] over two levels. First, aggregate votes are duly derived as inner quantifiers of the DMUs' consensual importance perceived from a preset optimism perspective. The aggregate votes are adopted, in the next level, as the order-inducing variable for calculating the ultimate efficiency scores. To substantiate the impact of subjectivity on the structure of the ranking patterns and, hence, assess the robustness of the proposed procedure, the IOWA operator is implemented with different optimism level values. The proposed methodology is applied for ranking a set of 25 material handling equipments (MHE) models collected from catalogues of MHE manufacturers and vendors.

2 Methodological Background

DEA is a non-parametric approach for evaluating DMUs' performance relative to an efficiency frontier. Conventional DEA models include CCR [22], and BCC [23]. For more on these models' development, see, e.g., [24] and [25]. A recent review on DEA can be found in [26].

2.1 Self- And Cross-Efficiency Models

Assume a set of K DMUs, each using M inputs w to yield N outputs z. The performance of each DMU d, identified with the values of its inputs w_{id} $(i = 1, \ldots, M)$ and outputs z_{jd} $(j = 1, \ldots, N)$, can be evaluated through its *self-efficiency* score e^*_{dd} or the *cross-efficiency* (CE) scores e^*_{pd} allocated by its peers p, for $p = 1, \ldots, K, p \neq d$.

The earliest DEA self-evaluation model is CCR, whose input oriented multiplier formulation is:

$$
\text{(CCR)} \quad
\begin{aligned}
& e_{dd}^{*} = \max \sum_{j=1}^{N} z_{jd}\mu_{jd} \\
& \text{Subject to:} \quad \sum_{i=1}^{M} w_{id}\nu_{id} = 1 \\
& \sum_{j=1}^{N} z_{jk}\mu_{jd} - \sum_{i=1}^{M} w_{ik}\nu_{id} \le 0 \quad k = 1, \ldots, K \\
& \mu_{jd} \ge 0 \qquad j = 1, \ldots, N \\
& \nu_{id} \ge 0 \quad i = 1, \ldots, M
\end{aligned}
$$

where e_{dd}^{*}, the self-efficiency score of DMU d, can be perceived as the optimal *appreciation* of d to itself [27]. Meanwhile, the associated multipliers μ_{jd}^{*} and ν_{id}^{*}, represent the weights selected by DMU d to quantity the influence of output z_{jd} and input w_{id}, respectively, on achieving its most advantageous self-evaluation [28].

Although model (CCR) can categorize DMU d as efficient ($e_{dd}^{*} = 1$) and inefficient ($e_{dd}^{*} < 1$), it may fail to achieve full ranking if several efficient DMUs occur [29]. We resort to the theory of DEA cross-efficiency (CE) [1] to cope with this situation.

Under a CE paradigm, each DMU d is permitted to evaluate its peers with its best own weight profile ($\boldsymbol{\mu}_{\boldsymbol{d}}^{*}\boldsymbol{\nu}_{\boldsymbol{d}}^{*}$). Yet, one may find different optimal solutions ($\boldsymbol{\mu}_{\boldsymbol{d}}^{*}\boldsymbol{\nu}_{\boldsymbol{d}}^{*}$) for the same objective e_{dd}^{*}, resulting in more than one CE score for the same CE evaluation and, hence, multiple ranking patterns. To palliate such a dearth, a number of alternative secondary goal models have been developed in the DEA literature (see, e.g., [30]), including the MRA model [31]. The prominence of the MRA model stems from its flexibility in allowing a specific DMU d to evaluate its peers with customized weight profiles, instead of the sole and unique set ($\boldsymbol{\mu}_{\boldsymbol{d}}^{*}\boldsymbol{\nu}_{\boldsymbol{d}}^{*}$) Consequently, discrimination is amplified early over the ranking process. The MRA model is formulated as follows.

$$
\text{(MRA)} \quad
\begin{aligned}
& e_{pd}^{*} = \max \sum_{j=1}^{N} z_{jp}\mu_{jp} \\
& \text{Subject to :} \quad \sum_{i=1}^{M} w_{ip}\nu_{ip} = 1 \\
& \sum_{j=1}^{N} z_{jk}\mu_{jp} - \sum_{i=1}^{M} w_{ik}\nu_{ip} \le 0 \quad k = 1, \ldots, K, k \ne p \\
& \sum_{j=1}^{N} z_{jp}\mu_{jp} - e_{dd}^{*} \sum_{i=1}^{M} w_{ip}\nu_{ip} = 0 \\
& \mu_{jp} \ge 0 \quad j = 1, \ldots, N \\
& \nu_{ip} \ge 0 \quad i = 1, \ldots, M
\end{aligned}
\qquad (1)
$$

e_{pd}^{*} is the CE score of DMU p, as assigned by the assessing DMU d, using the weight profile ($\boldsymbol{\mu}_{\boldsymbol{p}}^{*}\boldsymbol{\nu}_{\boldsymbol{p}}^{*}$) that maintains d's self-efficiency at its former level e_{dd}^{*}. Such a state is reinforced via constraint (1), which, simultaneously, reduces the feasibility space of CCR model and implies, with the maximization objective, that $e^*_{pd} \leq e^*_{dd}$ for all $p = 1, \ldots, K, p \neq d$.

In judgmental words, e^*_{pd} represents also the *level of appreciation* of d toward its peer p. Thus, it appears that, no matter the objectivity of the underlying DEA evaluation framework, each DMU's intrinsic value is only a reflection of an inner subjectivity. Subsequently, the efficiency scores e^*_{pd} and e^*_{dd}, $p = 1, \ldots, K$ and $d =$ 1, ..., K, are merely preferences under the most objective setting; DEA mathematical models.

Once the peer-evaluation of a DMU p is completed, it is presented with a set of CE scores e^*_{pd} The next step consists in using some aggregation procedure to amalgamate these scores, along e^*_{dd}, and produce p's ultimate efficiency score. Most CE approaches employ the arithmetic average as a device but the latter may fail to reflect the true performance of a DMU. Indeed, assigning equal weights to all efficiency scores means ignoring the relative importance of each score. To allow a more reasonable pondering of the relative importance of self- and CE scores from the viewpoint of the decision maker (DM), we use OWA [32] weights for the aggregation.

2.2 OWA Operator

The OWA operator is a multiple criteria aggregation tool that enables a reasonable incorporation of the DM's attitude throughout the decision process [33]. It is broadly applied in several disciplines, such as defects prioritization [34], selection of dispatching rules [35], and CE aggregation [19, 36, 37], to mention just a few areas.

An OWA operator with a weigh vector $\boldsymbol{\gamma} \in [0, 1]^K$ $[0, 1]^K$ is a function $g(\vartheta_{\boldsymbol{p}}; \boldsymbol{\gamma}) = \sum_{\ell=1}^{K} \gamma_\ell \vartheta_{p\ell}$ where $\vartheta_{p\ell}$ is the value of the ℓth factor of the argument $\vartheta_{\boldsymbol{p}} = (\vartheta_{p1}\vartheta_{p2}...\vartheta_{pK})$ and γ_ℓ is the associated OWA weight, with $\sum_{\ell=1}^{K} \gamma_\ell = 1$.

The OWA weight vector $\boldsymbol{\gamma}$ can be generated in a way that echoes the subjectivity level of the DM. Although there are several approaches for generating these weights (see, e.g., [38]), the minimax disparity models are the most frequently used approaches. In our study, the following model, due to Wang and Parkan [39], is used.

$$
\text{(WP)} \quad \begin{array}{l} \min\ \delta \\ s.t. \\ \sum_{\ell=1}^{K-1} \left(\frac{K-\ell}{K-1}\right) \gamma_\ell = \alpha \quad \alpha \in [0, 1] \end{array} \tag{2}
$$

$$
\sum_{\ell=1}^{K} \gamma_\ell = 1 \tag{3}
$$

$$
\begin{array}{ll}
-\delta \le \gamma_\ell - \gamma_{\ell+1} \le \delta & \ell = 1, \ldots, K-1 \\
\gamma_\ell \ge 0 & \ell = 1, \ldots, K
\end{array}
\tag{4}
$$

Model (WP) aims at minimizing the deviation δ between successive aggregation weights γ_ℓ and $\gamma_{\ell+1}$, $\ell = 1, ..., K$, as formulated by the set of constraints (4). The parameter α on the right hand side of constraint (2), represents the level of optimism of the DM, also known as orness value [33]. The extreme values of α are $\alpha = 0$ and $\alpha = 1$, which correspond to purely pessimistic and purely optimistic DMs, respectively. A neutral attitude is quantified with $\alpha = 0.5$. Thus, orness values $0.5 < \alpha < 1$ reflect DMs that are just optimistic. Optimism being a subjective stance, the variability of α offers an opportunity to incorporate a broader range of DM's subjectivity levels over the ranking process in a regulated manner.

3 A Two-Level Induced OWA Procedure for Ranking

The DEA-based methodology that we propose deploys as illustrated in Fig. 1.

As explained in the methodological background, self and cross-evaluation systems may involve CCR and MRA models to produce the CE matrix, whereas subjectivity regulation refers to the choice of an appropriate OWA model to generate the aggregation weights under a specified optimism level α. The preference-voting system as well as the two IOWA levels will be detailed in the next sections.

3.1 Preference Voting

Once the cross evaluation is completed, the resulting CE matrix $\boldsymbol{E}$ for $d = 1, ..., K$, is as follows

$$
\boldsymbol{E} = \begin{bmatrix} e^*_{11} & e^*_{12} & \ldots & e^*_{1K} \\ e^*_{21} & e^*_{22} & \ldots & e^*_{2K} \\ \ldots & \ldots & \ldots & \ldots \\ e^*_{K1} & e^*_{K2} & \ldots & e^*_{KK} \end{bmatrix}
$$

Column $\boldsymbol{e}^*_p$ in matrix $\boldsymbol{E}$ holds the CE scores of DMU p as assigned by all assessing DMUs d ($d = 1, \ldots, K$). Thus, $\boldsymbol{e}^*_p$ symbolizes the collective evaluation of DMU p for $p = 1, \ldots, K$. Thus, the CE matrix $\boldsymbol{E}$ can be implicitly viewed as a voting framework where each DMU holds a dual status of candidate and constituent, and deemed free to vote without outer influence [39]. The strength of such an approach resides in the collective consensus among DMUs [36] that the associated preference-voting system aptly reflects.

Cross ranking entails associating to the CE matrix $\boldsymbol{E}$ a preference voting matrix Θ through the following steps:

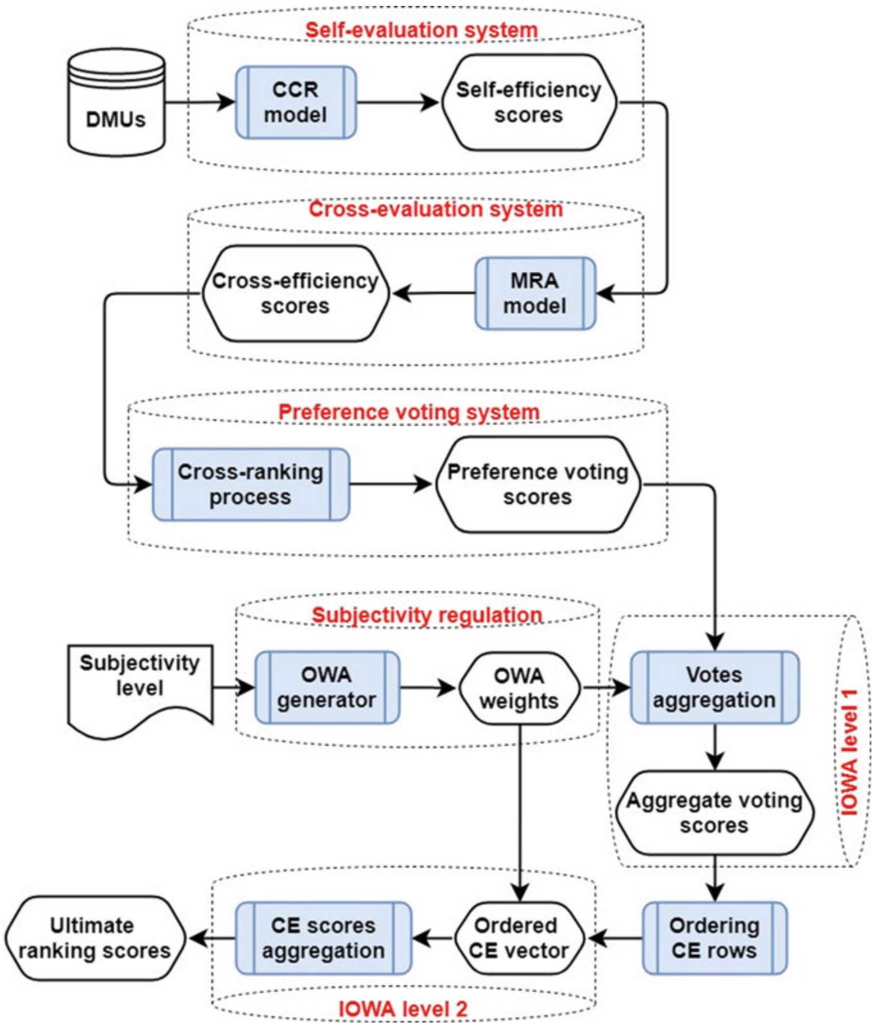


Fig. 1 A two-level induced OWA ranking procedure

Step 1: Sort separately each row-vector of CE scores $\boldsymbol{e}_{\boldsymbol{d}}^{*} = (e_{d1}^{*}e_{d2}^{*}...e_{dK}^{*})$ in descending order to determine the preference of the assessing DMU d towards each peer p. The higher the value of e_{dp}^{*} the more important is p from d's perspective.

Step 2: Assign to each score e_{dp}^{*}, $p = 1, ..., K$, the corresponding rank position ρ_{dp}. This leads to the cross-ranking matrix $\boldsymbol{R}$ where

$$\boldsymbol{R} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1K} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{K1} & \rho_{K2} & \cdots & \rho_{Kk} \end{bmatrix}$$

Each rank position ρ_{dp}. can take a value f, with $1 \leq f \leq K$, and f can occur more than once in the same column p, indicating that DMU p has been ranked fth by more than one assessing DMU d.

Step 3: Count ϑ_{pf} the number of DMUs that ranked DMU p at the fth position and build the preference voting matrix Θ, where

$$\Theta = \begin{bmatrix} \vartheta_{11}\, \vartheta_{12} \ldots \vartheta_{1K} \\ \vartheta_{21}\, \vartheta_{22} \ldots \vartheta_{2K} \\ \ldots \ldots \ldots \ldots \\ \vartheta_{K1}\, \vartheta_{K2} \ldots \vartheta_{KK} \end{bmatrix}$$

Each DMU p is associated with a preference voting vector $\vartheta_{\boldsymbol{p}} = (\vartheta_{p1}\vartheta_{p2}\ldots\vartheta_{pK})$, where ϑ_{pf} is the number of votes in support of ranking DMU p at the fth position, with $K = \sum_{f=1}^{K} \vartheta_{pf}$ voters. Thus, matrix Θ encompasses the importance of each DR from a cooperative stance, entailing all DMUs. As such, the importance scale that Θ reflects is much more consensual.

With the importance scale determined via Θ, the ranking of the DMUs can still not be performed without a priori ranking scores. Intuitively, the average voting score would appear as a valid metric; an option that is inevitably discarded because the total number of votes in each row p is the same i.e., $K = \sum_{f=1}^{K} \vartheta_{pf}$, leading to equal averages. However, knowing that the importance of the votes ϑ_{pf} over each row p is already established through the rank orders, the Induced OWA operator [40] might be an appropriate ground for the aggregation process, as shown in the following section.

3.2 *Induced OWA Aggregation*

In OWA operator, the arguments are reordered based on their magnitudes, prior to the aggregation. The induced OWA operator (IOWA) [40] is an important extension of OWA operator where the arguments' ordering is merely induced by order-inducing variables [41, 42]. The order-inducing variable can be defined depending on the type of information used for ordering, which may involve importance, consistency, preference, or any other type.

Induced OWA level 1

In regard to the CE framework, the preference voting system embedded under the CE matrix $\boldsymbol{E}$ reflects the inner importance of the DMUs, which remains consistent with the observed values of the inputs and outputs as well as the secondary goal DEA model adopted to compute $\boldsymbol{E}$. As such, the preference-voting matrix Θ can be legitimately considered as an order-inducing variable for the first IOWA aggregation level (see Fig. 1). Meanwhile, the DM's attitude towards the arguments of the aggregation process is measured via his/her optimism level, which expresses the outer preference. As opposed to the former preference, the latter can be regulated through diversifying the values of the optimism level α. For each specified α, model (WP) is solved to generate a vector $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{K-1}, \gamma_K)$ of OWA weights. Given a vector of votes $\vartheta_{\boldsymbol{p}} = (\vartheta_{p1}\vartheta_{p2}...\vartheta_{pK})$ related to DMU p, using vector $\boldsymbol{\gamma}$ as an aggregation device enables the relative importance of the votes to be realistically weighted from a collective standpoint. Thus, the aggregate vote υ_p associated to DMU p can be calculated as:

$$\upsilon_p = \sum_{\ell=1}^{K} \gamma_\ell \vartheta_{p\ell}. \tag{5}$$

High values of υ_p imply necessarily that a high number of DMUs voted for DMU p to be in leading rank positions regardless of the magnitude of the CE score that was attached to each vote.

Sorting the elements of vector $\boldsymbol{\upsilon} = (\upsilon_1, ..., \upsilon_K)$ from the highest to the lowest may yield a ranking pattern that is sufficiently robust, yet, ignoring the CE score context may affect the reliability of the results. Thus, we propose a second aggregation level that restores the CE scores as a basis for the ultimate ranking process.

Induced OWA level 2

The vector of aggregate votes $\boldsymbol{\upsilon} = (\upsilon_1, ..., \upsilon_K)$ can be viewed as an order-inducing variable for the second IOWA aggregation level. At this level, given that each row of matrix $\boldsymbol{E}$ is associated with an aggregate vote υ_p, a sorted vector $\boldsymbol{\upsilon}$ can be used as a device for reordering the rows of $\boldsymbol{E}$ from the best consensual perspective. Let $\boldsymbol{E}^{\mathbf{O}}$ denote the reordered CE matrix $\boldsymbol{E}$

$$\boldsymbol{E^o} = \begin{bmatrix} e^o_{11} & e^o_{12} & \cdots & e^o_{1K} \\ e^o_{21} & e^o_{22} & \cdots & e^o_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ e^o_{K1} & e^o_{K2} & \cdots & e^o_{KK} \end{bmatrix}$$

The ultimate efficiency score ϕ_p. can be calculated for each DMU p, $p = 1, \ldots, K$, by an OWA aggregation on the corresponding column $\boldsymbol{e^o_p}$ as:

$$\phi_p = \sum_{\ell=1}^{K} \gamma_\ell e^o_{p\ell}. \quad (6)$$

ϕ_p. can be viewed as a fusion of the inner and outer preferences over the ranking process.

4 Illustrative Examples

To illustrate the proposed approach, we use the numerical example in [43], which consists of 12 manufacturing systems S_k, $k = 1, \ldots, 12$, described with 2 inputs to produce 4 outputs. The data sample is provided in Table 14 in the Appendices.

All computations relating to LP optimization were carried out using IBM-ILOG CPLEX version 12.6.4, which was embedded under the C++ code of the ranking algorithm. The optimal solutions of the CCR model are provided in Table 1.

As bold highlighted in Table 1, 7 out of 12 systems are identified as efficient and, hence, self-efficiency fails to achieve full discrimination among the manufacturing systems. To achieve full ranking, we implement the proposed procedure, which starts by solving model (MRA) 11 times for each DMU p, $p = 1, \ldots, 12$. The corresponding CE matrix $\boldsymbol{E}$ is produced in Table 2.

Table 1 Self-efficiency scores for 12 DMUs

DMU	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
e^*_{dd}	**1**	**1**	0.982	**1**	**1**	**1**	**1**	0.961	**1**	0.954	0.983	0.801

Table 2 Peer evaluation scores for 12 DMUs

DMU	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
S_1	1	1	0.982	1	1	0.977	1	0.961	0.797	0.857	0.983	0.801
S_2	1	1	0.890	1	1	0.851	0.882	0.911	0.644	0.736	0.923	0.780
S_3	1	0.959	0.982	0.927	1	0.932	1	0.927	0.422	0.760	0.977	0.801
S_4	1	1	0.953	1	1	1	1	0.961	0.996	0.915	0.953	0.801
S_5	1	1	0.982	1	1	0.995	1	0.961	0.885	0.954	0.977	0.801
S_6	0.915	0.920	0.904	1	0.988	1	1	0.919	1	0.862	0.865	0.758
S_7	1	0.990	0.982	1	1	1	1	0.961	1	0.954	0.983	0.801
S_8	1	0.977	0.949	1	1	0.962	1	0.961	0.753	0.833	0.951	0.794
S_9	0.831	0.815	0.839	0.995	0.916	1	1	0.850	1	0.919	0.818	0.701
S_{10}	0.764	0.704	0.904	0.956	1	0.951	1	0.860	0.849	0.954	0.714	0.720
S_{11}	1	0.966	0.924	0.896	0.927	0.954	1	0.945	0.672	0.783	0.983	0.759
S_{12}	1	0.984	0.953	1	1	0.967	1	0.950	0.724	0.795	0.953	0.801

The next step consists in cross-ranking the systems in the light of the CE scores that are assigned. The cross- ranking matrix $\boldsymbol{R}$ is as presented in Table 3.

For instance, $\rho_{16} = 10$ means that S_1 ranked S_6 at the 10th position, while nine systems were ranked 1st. By examining column S_6, we can count the number of votes ϑ_{6f} in favour of ranking S_6 at the fth position. Thus, it appears that $\vartheta_{6\,10} = 6$ systems ranked S_6 at the 10^{th} position, $\vartheta_{61} = 4$ systems ranked S_6 at the 1st position, and $\vartheta_{66} = \vartheta_{69} = 1$ indicates that only one system ranked S_6 at the 6th and 9th positions. As a result, the preference-voting matrix Θ, shown in Table 4, is drawn by following the same reasoning for each system,

Table 3 Cross ranking matrix $\boldsymbol{R}$ for 12 DMUs

DMU	S_1	S_2		S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
S_1	1	1		1	1	1	10	1	1	11	12	1	1
S_2	1	1		9	1	1	10	5	7	11	12	8	6
S_3	1	11		4	5	1	10	1	7	12	9	8	6
S_4	1	1		11	1	1	1	1	1	9	10	12	1
S_5	1	1		9	1	1	10	1	1	12	1	11	1
S_6	6	12		11	1	5	1	1	8	1	10	9	7
S_7	1	12		1	1	1	1	1	1	1	1	1	1
S_8	1	10		8	1	1	9	1	5	12	11	7	6
S_9	7	11		12	4	5	1	1	8	1	6	10	9
S_{10}	7	12		11	5	1	6	1	8	4	1	10	9
S_{11}	2	9		4	7	4	10	2	8	11	12	1	6
S_{12}	1	8		6	1	1	10	1	7	12	11	9	1

Table 4 Preference-voting matrix Θ for 12 DMUs

DMU	1	2	3	4	5	6	7	8	9	10	11	12
S_1	8	1	0	0	0	1	2	0	0	0	0	0
S_2	4	0	0	0	0	0	0	1	1	1	2	3
S_3	2	0	0	2	0	1	0	1	2	0	3	1
S_4	8	0	0	1	2	0	1	0	0	0	0	0
S_5	9	0	0	1	2	0	0	0	0	0	0	0
S_6	4	0	0	0	0	1	0	0	1	6	0	0
S_7	10	1	0	0	1	0	0	0	0	0	0	0
S_8	4	0	0	0	1	0	3	4	0	0	0	0
S_9	3	0	0	1	0	0	0	0	1	0	3	4
S_{10}	3	0	0	0	0	1	0	0	1	2	2	3
S_{11}	3	0	0	0	0	0	1	2	2	2	1	1
S_{12}	5	0	0	0	0	4	1	0	2	0	0	0

In Table 4, the columns represent possible rank positions of each DMU. Row S_7 reveals that 10 out of 12 DMUs voted in favour of S_7 to be first, suggesting that S_7 is likely to be the leading system, followed probably by S_5 which collected 9 votes for the first rank position, along with a single vote for the 4th position and two votes for the 5th position. Obviously, the picture is now better than the one reflected through self-efficiency. Even here, although holders of leading positions can be foreseen for some data samples, the picture is generally fuzzier, with most DMUs presenting blends of votes that render any speculation almost impossible.

Noting that the total number of votes in each row p is $K = 12$, for $p = 1, \ldots, 12$, the average voting score is excluded as an option for ranking the DMUs. However, it is always possible to exploit the importance scale prompted from the preference-voting matrix Θ and apply an OWA operator to derive an aggregate vote υ_p for each row p.

We start by solving model (WP) for $K = 12$ to generate a vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{12})$ of OWA weights. We use a broad range of DM's subjectivity levels $\alpha \in \{0.55, 0.60, 0.65, 0.69\}$ as a way to adequately appraise the robustness of the ranking process through diversifying the decision contexts. The OWA weights produced for each value α are displayed in Table 5.

Using formula (5), we compute the vectors of aggregate votes $\boldsymbol{\upsilon}$ presented in Table 6.

For example, the aggregate weights in row S_1 column $\alpha = 0.55$ is calculated by multiplying column $\alpha = 0.55$ in Table 5 by row S_1 in Table 4, so that $\upsilon_1 = \sum_{\ell=1}^{12} \gamma_\ell \vartheta_{1\ell} = 0.10449 \times 8 + 0.10064 \times 1 + 0.09679 \times 0 + \ldots + 0.06218 \times 0 = 1.18462$.

Note that the votes do not require a reordering prior to the aggregation since the order is already established via the voting system.

Table 5 OWA weights generated for different subjectivity levels α with K=12

OWA	$\alpha = 0.55$	$\alpha = 0.60$	$\alpha = 0.65$	$\alpha = 0.69$
γ_1	0.10449	0.12564	0.14679	0.16372
γ_2	0.10064	0.11795	0.13526	0.14910
γ_3	0.09679	0.11026	0.12372	0.13449
γ_4	0.09295	0.10256	0.11218	0.11987
γ_5	0.08910	0.09487	0.10064	0.10526
γ_6	0.08526	0.08718	0.08910	0.09064
γ_7	0.08141	0.07949	0.07756	0.07603
γ_8	0.07756	0.07179	0.06603	0.06141
γ_9	0.07372	0.06410	0.05449	0.04679
γ_{10}	0.06987	0.05641	0.04295	0.03218
γ_{11}	0.06603	0.04872	0.03141	0.01756
γ_{12}	0.06218	0.04103	0.01987	0.00295

Table 6 Aggregate votes and associate row-ranks for different subjectivity levels α

DMU	α = 0.55		α = 0.60		α = 0.65		α = 0.69	
	υ_p	r_p	υ_p	r_p	υ_p	r_p	υ_p	r_p
S_1	1.18462	4	1.36923	4	1.55385	4	1.70154	4
S_2	0.95769	10	0.91538	10	0.87308	10	0.83923	10
S_3	0.96538	8	0.93077	9	0.89615	8	0.86846	8
S_4	1.18846	3	1.37692	3	1.56538	3	1.71615	3
S_5	1.21154	2	1.42308	2	1.63462	2	1.80385	2
S_6	0.99615	7	0.99231	7	0.98846	7	0.98538	7
S_7	1.23462	1	1.46923	1	1.70385	1	1.89154	1
S_8	1.06154	6	1.12308	6	1.18462	6	1.23385	6
S_9	0.92692	12	0.85385	12	0.78077	12	0.72231	12
S_{10}	0.93077	11	0.86154	11	0.79231	11	0.73692	11
S_{11}	0.96538	9	0.93077	8	0.89615	9	0.86846	9
S_{12}	1.09231	5	1.18462	5	1.27692	5	1.35077	5

From the aggregate votes υ_p, we work out the associated ranks r_p presented along υ_p in Table 6. The ranks r_p are needed for reordering the rows of CE matrix $\boldsymbol{E}$ in the second IOWA aggregation level. Based on the resulting ranks, the structure of $\boldsymbol{E^O}$ is the same for α = 0.65 and α = 0.69 but slightly different for α = 0.60. Accordingly, the reordered CE matrix $\boldsymbol{E^O}$ for α = 0.55 is as shown in Table 7.

The advantage of $\boldsymbol{E^O}$ resides in the fact that it standardizes the order of the CE scores for all the columns. Instead of sorting the elements of each column separately, based solely on their magnitudes (traditional OWA procedure), the order is induced

Table 7 Reordered CE matrix $\boldsymbol{E^O}$ for α = 0.55

DMU	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
S_7	1	0.990	0.982	1	1	1	1	0.961	1	0.954	0.983	0.801
S_5	1	1	0.982	1	1	0.995	1	0.961	0.885	0.954	0.977	0.801
S_4	1	1	0.953	1	1	1	1	0.961	0.996	0.915	0.953	0.801
S_1	1	1	0.982	1	1	0.977	1	0.961	0.797	0.857	0.983	0.801
S_{12}	1	0.984	0.953	1	1	0.967	1	0.950	0.724	0.795	0.953	0.801
S_8	1	0.977	0.949	1	1	0.962	1	0.961	0.753	0.833	0.951	0.794
S_6	0.915	0.920	0.904	1	0.988	1	1	0.919	1	0.862	0.865	0.758
S_3	1	0.959	0.982	0.927	1	0.932	1	0.927	0.422	0.760	0.977	0.801
S_{11}	1	0.966	0.924	0.896	0.927	0.954	1	0.945	0.672	0.783	0.983	0.759
S_2	1	1	0.890	1	1	0.851	0.882	0.911	0.644	0.736	0.923	0.780
S_{10}	0.764	0.704	0.904	0.956	1	0.951	1	0.860	0.849	0.954	0.714	0.720
S_9	0.831	0.815	0.839	0.995	0.916	1	1	0.850	1	0.919	0.818	0.701

Table 8 Ultimate efficiency scores and associated DMUs' ranks for 12 manufacturing systems

DMU	α = 0.55		α = 0.60		α = 0.65		α = 0.69	
	ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p
S_1	**0.96702**	**5**	**0.97484**	**4**	**0.98266**	**4**	**0.98891**	**4**
S_2	**0.95241**	**6**	**0.96187**	**6**	**0.97122**	**6**	**0.97874**	**5**
S_3	0.94281	7	0.94807	7	0.95423	7	0.95880	7
S_4	**0.98347**	**3**	**0.98552**	**3**	**0.98804**	**3**	**0.98987**	**3**
S_5	**0.98841**	**2**	**0.99035**	**2**	**0.99342**	**2**	**0.99542**	**2**
S_6	**0.96896**	**4**	**0.97226**	**5**	**0.97521**	**5**	**0.97771**	**6**
S_7	**0.99175**	**1**	**0.99334**	**1**	**0.99493**	**1**	**0.99620**	**1**
S_8	0.93597	8	0.94128	8	0.94631	9	0.95045	9
S_9	0.81964	11	0.82940	11	0.83531	11	0.84158	11
S_{10}	0.86416	10	0.86834	10	0.87217	10	0.87538	10
S_{11}	0.93169	9	0.94020	9	0.94862	8	0.95539	8
S_{12}	0.78085	12	0.78483	12	0.78944	12	0.79288	12

from the aggregate votes and applied to the whole row. Such a strategy aims to enhance the consistency of the aggregation process. Therefore, the ultimate efficiency scores ϕ_p are calculated for each column of $\boldsymbol{E^O}$ using formula (6) and the resulting values and associates ranks ur_p are exhibited in Table 8.

Here, if we consider ϕ_3 for α = 0.55, it is computed by multiplying the column-vector of OWA weights in Table 5 by column $\boldsymbol{e}^3$ in Table 7 so that $\phi_3 = \sum_{\ell=1}^{12} \gamma_\ell e^o_{3\ell} = 0.10449 \times 0.982 + 0.10064 \times 0.982 + \ldots + 0.06218 \times 0.839 = 0.94281$.

Regardless of the subjectivity level α, the ranking patterns identify S_7 and S_{12} as the best and the worst manufacturing systems, respectively. Generally, the DMUs preserve the same rank positions with only a few position swaps, where the rank gap does not exceed one in the worst case. The fact that the ranking patterns exhibit the same structure for all subjectivity levels may also indicate that the variability of the DM's attitude does not affect the ranking process. Moreover, it is manifest that the six leading manufacturing systems are S_1, S_2, S_4, S_5, S_6 and S_7 are all strongly efficient DMUs, as bolded in Table 8. This supports the effectiveness of the proposed approach in enhancing discrimination since the occurrence of multiple efficient DMUs is the main issue with self-efficiency. All these ingredients testify in favor of the robustness of the proposed ranking procedure.

In order to ascertain the robustness of the proposed ranking procedure, we use two additional examples from the literature. The ranking algorithm is implemented for each new set of DMUs under decision making conditions that involve different subjectivity levels $\alpha \in \{0.55, 0.60, 0.65, 0.69\}$.

The first additional example is drawn from [44] and comprises 14 bank branches, each defined with 3 inputs and 4 outputs The full set of data is available in Table 15 in the Appendices. Table 9 presents the results for this example.

Table 9 Ultimate efficiency scores and associated DMUs' ranks for 14 bank branches

DMU	e^*_{pp}	α = 0.55		α = 0.60		α = 0.65		α = 0.69	
		ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p
B_1	**1**	**0.67338**	**13**	**0.68048**	**13**	**0.68759**	**13**	**0.69327**	**13**
B_2	**1**	**0.98922**	**2**	**0.98861**	**2**	**0.98800**	**2**	**0.98751**	**2**
B_3	**1**	**0.95909**	**3**	**0.96646**	**3**	**0.97382**	**3**	**0.97971**	**3**
B_4	**1**	**0.87059**	**8**	**0.87767**	**8**	**0.88475**	**8**	**0.89041**	**8**
B_5	0.90411	0.78806	10	0.79278	10	0.79749	10	0.80126	10
B_6	**1**	**0.99157**	**1**	**0.99442**	**1**	**0.99727**	**1**	**0.99955**	**1**
B_7	0.78209	0.73808	12	0.74419	12	0.75031	12	0.75520	12
B_8	**1**	**0.93685**	**4**	**0.94687**	**4**	**0.95689**	**4**	**0.96490**	**4**
B_9	**1**	**0.53205**	**14**	**0.56218**	**14**	**0.59232**	**14**	**0.61642**	**14**
B_{10}	**1**	**0.89569**	**5**	**0.91262**	**5**	**0.92956**	**5**	**0.94311**	**5**
B_{11}	0.96685	0.87556	7	0.88719	7	0.89882	7	0.90812	7
B_{12}	0.85225	0.79251	9	0.80239	9	0.81226	9	0.82016	9
B_{13}	0.90491	0.74935	11	0.75883	11	0.76831	11	0.77589	11
B_{14}	**1**	**0.89535**	**6**	**0.90635**	**6**	**0.91735**	**6**	**0.92615**	**6**

The bold highlighted rows in Table 9 refer to the efficient bank branches. It is visible that seven out of nine efficient banks are taking the lead, namely, B_2, B_3, B_4, B_6, B_8, B_{10}, and B_{14}, which advocates again in favour of the discriminating potential of the new ranking approach. Moreover, the ranking patterns are ostensibly similar no matters the DM's optimism level.

The data sample of the second example is the same used in [45], which involves 14 airlines, defined with 3 inputs and 2 outputs (see Table 16 in the Appendices). The associated ultimate efficiency scores ϕ_p and related ranking patterns are reproduced in Table 10.

A same stance is further established through the results shown in Table 10 where all the efficient DMUs (bold highlighted rows in Table 10) are found on the top of the list besides the steadiness of the ranking patterns, which preserve the overall structure across different optimism levels in spite of slight rank position swapping between A_7 and A_{10} for $\alpha = 0.69$.

Consequently, with these supporting examples, the conclusion on the robustness of the ranking procedure is more corroborated.

5 Application

The proposed procedure is applied for selecting the best material handling equipment (MHE) model to purchase. We use a sample of 25 MHE, collected from manufacturers

Table 10 Ultimate efficiency scores and associated DMUs' ranks for 14 airlines

DMU	e^*_{pp}	α = 0.55		α = 0.60		α = 0.65		α = 0.69	
		ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p
A_1	0.86836	0.77251	12	0.79069	12	0.80886	12	0.82248	12
A_2	0.33794	0.23164	14	0.23666	14	0.24169	14	0.24580	14
A_3	0.94752	0.80379	10	0.82523	10	0.84668	10	0.86317	10
A_4	0.95809	0.84210	8	0.86201	8	0.88193	8	0.89692	8
A_5	**1**	**0.91547**	**4**	**0.93234**	**4**	**0.94921**	**4**	**0.96316**	**4**
A_6	0.97658	0.79536	11	0.81569	11	0.83603	11	0.85123	11
A_7	**1**	**0.85832**	**6**	**0.87711**	**6**	**0.89590**	**6**	**0.90919**	**7**
A_8	0.85876	0.76098	13	0.77693	13	0.79289	13	0.80525	13
A_9	0.94774	0.82137	9	0.83910	9	0.85684	9	0.87147	9
A_{10}	**1**	**0.85159**	**7**	**0.87305**	**7**	**0.89450**	**7**	**0.91390**	**6**
A_{11}	**1**	**0.93544**	**2**	**0.95155**	**2**	**0.96767**	**2**	**0.98057**	**2**
A_{12}	**1**	**0.91576**	**3**	**0.93477**	**3**	**0.95378**	**3**	**0.96964**	**3**
A_{13}	**1**	**0.93931**	**1**	**0.95853**	**1**	**0.97775**	**1**	**0.99313**	**1**
A_{14}	**1**	**0.88545**	**5**	**0.90471**	**5**	**0.92397**	**5**	**0.93851**	**5**

and vendors' catalogues, where each MHE is described with 2 inputs and 2 outputs, as shown in Table 11.

The input variables include the total cost of MHE and the floor space it uses. The total cost comprises the costs to purchase and setup the equipment and its components. For example, it is the purchasing cost of the AGV, its track, software, and battery charger, as well as the cost of installing the track and testing the AGV system. The floor space refers to the area occupied by the equipment or used for its movement.

The output variables are the weight capacity, the salvage value and the flexibility of a MHE. The weight capacity allows supporting and carrying loads, and is a common attribute among transfer, storage, and unit load equipment. The salvage value is the estimated resale value of an MHE at the end of its life-cycle. Flexibility expresses the sum of actions and features of MHE in performing operations, estimated as shown in Table 12.

The new ranking algorithm is run with the proposed data set. The CE matrix $\boldsymbol{E}$, the preference-voting matrix Θ, and the OWA weights produced over the application case are given in Tables 17, 18 and 19, respectively, in the Appendices.

Using formula (5), we compute the vectors of ultimate efficiency scores ϕ_p presented in Table 13, and we work out the associated ranks ur_p.

Irrespective of the optimistic level α, there is no doubt that the leading MHE is m_{25}, which swap position with m_{07}, the next top MHE, for the optimism level $\alpha = 0.68$. Moreover, 5 out of 6 MHE that are on top of the list, specifically m_{25}, m_{07}, m_{02}, m_{04}, m_{05} and m_{23}, have been classified strongly efficient in the self-efficiency evaluation, as bold highlighted in Table 13. This confirms again the contribution of

Table 11 Data sample for MHE selection

ID		Inputs		Outputs	
		Floor space (ft^2)	Cost ($)	Capacity (Ib)	Flexibility
m_{01}	R. conveyor Relius	108.3	10,885	25,000	4
m_{02}	R. conveyor Reno	80	8288	17,600	4
m_{03}	R. conveyor Hytrol	86.7	8700	18,000	4
m_{04}	R. conveyor TGW	102	9250	30,000	4
m_{05}	B. conveyor Douglas	125	8700	27,000	4
m_{06}	B. conveyor Reno	112.5	11,060	25,000	4
m_{07}	B. conveyor Conveyor direct	81.7	7640	18,000	4
m_{08}	Swing reach truck Raymond	170	37,000	5500	6
m_{09}	Swing reach truck Drexel	140	39,000	6000	6
m_{10}	Swing reach truck Crown	150	37,500	5000	6
m_{11}	Swing reach truck Bendi	195	45,000	6000	6
m_{12}	Lift truck Bendi B3	180	24,000	5500	5
m_{13}	Lift truck Bendi B4	195	19,000	5000	5
m_{14}	Lift truck CLARK	233	18,500	5000	5
m_{15}	Lift truck Hyster	161	22,000	6000	5
m_{16}	Lift truck Toyota	179	25,000	5000	5
m_{17}	AGV Corecon R301	163	200,000	5000	5
m_{18}	AGV Falcon F150	121	262,500	6000	5
m_{19}	AGV Amerden FLA2200	108	240,000	6000	5
m_{20}	AGV Amerden FLB2000	113	210,000	6000	5
m_{21}	AGV JBT	108	195,000	5000	5
m_{22}	VLM Hanel	95	69,350	132,450	7
m_{23}	VLM Kardex	91	60,900	264,550	7
m_{24}	VLM Vidmar MAV 75D	99	63,750	132,000	7
m_{25}	VLM Vidmar MCV 75D	78	51,500	150,000	7

Table 12 Flexibility matrix of MHEs

Action and Feature	R conveyor	B conveyor	Swing reach truck	Lift truck	AGV	VLM
Transfer	•	•	•	•	•	
Lift			•	•	•	•
Load			•	•	•	•
Unload			•	•	•	•
Store						•
Variable path			•	•		
Overhead/ V. space use	•	•				•
No labor for load transfer	•	•			•	•
Aisle space saving			•			
Handling and processing	•	•				
Item security/protection						•
Flexibility	4	4	6	5	5	7

the proposed ranking approach to boosting discrimination while keeping consist the outcomes.

On another hand, the ranking patterns present slight dissimilarities, observed only for the top 10 MHEs, as shown in Fig. 2.

While m_{02}, m_{03}, m_{04} and m_{23} preserve the same ranks no matters the value of α, the pairs (m_{07}, m_{25}), (m_{05}, m_{24}),(m_{07}, m_{25}) and (m_{01}, m_{22}), swap positions at α = 0.68. The MHE m_{05} and m_{22} appear as the only MHE that shift position at α = 0.65, exhibiting a maximum gap of 2 rank positions between α = 0.60 and α = 0.68, which remains practically acceptable. In practice, such "tight" dissimilarities may be perceived as more alternatives to select amongst if the market circumstances are less favorable to the ideal alternative.

In the light of the above discussion, the procurement manager can opt to purchase m_{25} or m_{07}, depending on the operating as well as the market conditions, which may be suitably quantified via an optimism level α.

6 Conclusion

In this paper, we discussed a new approach for ranking DMUs under a CE evaluation framework. The novelty of the proposed approach resides in its handling of the CE aggregation process under a two-level Induced OWA procedure that exploits ingeniously the preference-voting system embedded under the CE matrix. Such approach

Table 13 Ranking patterns of MHEs for different subjectivity levels α

MHE	e^*_{pp}	α = 0.55		α = 0.60		α = 0.65		α = 0.68	
		ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p	ϕ_p	ur_p
m_{01}	0.80827	0.68377	10	0.68937	10	0.69491	10	0.69801	9
m_{02}	**1**	**0.88514**	**3**	**0.89009**	**3**	**0.89505**	**3**	**0.89802**	**3**
m_{03}	0.92958	0.82592	5	0.83099	5	0.83605	5	0.83907	5
m_{04}	**1**	**0.77369**	**6**	**0.78207**	**6**	**0.79026**	**6**	**0.79451**	**6**
m_{05}	**1**	**0.69077**	**9**	**0.69980**	**9**	**0.70852**	**8**	**0.71266**	**7**
m_{06}	0.79078	0.66348	11	0.66909	11	0.67464	11	0.67772	11
m_{07}	**1**	**0.89512**	**2**	**0.90141**	**2**	**0.90769**	**2**	**0.91144**	**1**
m_{08}	0.60703	0.47850	15	0.47551	15	0.47258	15	0.47094	15
m_{09}	0.68580	0.53652	12	0.53200	12	0.52759	12	0.52512	12
m_{10}	0.66168	0.51829	13	0.51435	13	0.51050	13	0.50835	13
m_{11}	0.52081	0.40982	19	0.40706	19	0.40437	19	0.40287	19
m_{12}	0.53293	0.43663	17	0.43611	17	0.43561	17	0.43534	17
m_{13}	0.51943	0.43915	16	0.44008	16	0.44101	16	0.44156	16
m_{14}	0.51622	0.39249	20	0.39483	20	0.39710	20	0.39823	20
m_{15}	0.59313	0.48541	14	0.48476	14	0.48414	14	0.48380	14
m_{16}	0.53130	0.43276	18	0.43198	18	0.43123	18	0.43081	18
m_{17}	0.34181	0.21682	25	0.21077	25	0.20481	25	0.20140	25
m_{18}	0.46045	0.23564	24	0.22614	24	0.21673	24	0.21126	24
m_{19}	0.51587	0.26175	23	0.25105	23	0.24046	23	0.23430	23
m_{20}	0.49305	0.26696	22	0.25712	22	0.24741	22	0.24177	22
m_{21}	0.51587	0.28210	21	0.27187	21	0.26177	21	0.25592	21
m_{22}	0.82105	0.71092	8	0.70483	8	0.69889	9	0.69541	10
m_{23}	**1**	**0.87056**	**4**	**0.87236**	**4**	**0.87416**	**4**	**0.87474**	**4**
m_{24}	0.79824	0.72066	7	0.71629	7	0.71206	7	0.70958	8
m_{25}	**1**	**0.91809**	**1**	**0.91335**	**1**	**0.90876**	**1**	**0.90601**	**2**

throws more objectivity on the ranking process as it does not necessitate from the DM to set his/her preferences a priori.

We used a data sample of 12 manufacturing systems to illustrate different steps of the ranking mechanism. The latter sample has been used, along two other samples of DMUs drawn from the literature, to investigate the robustness of the patterns produced over the ranking process. The algorithm has been deployed using different operating conditions, reflected specifically through a range of optimism levels of the DM. The results indicate primarily that ranking with the IOWA aggregation approach leads to ranking patterns that are robust enough.

The proposed procedure has been applied to address the problem of selecting the best MHE model to purchase from a set of available options in the market.

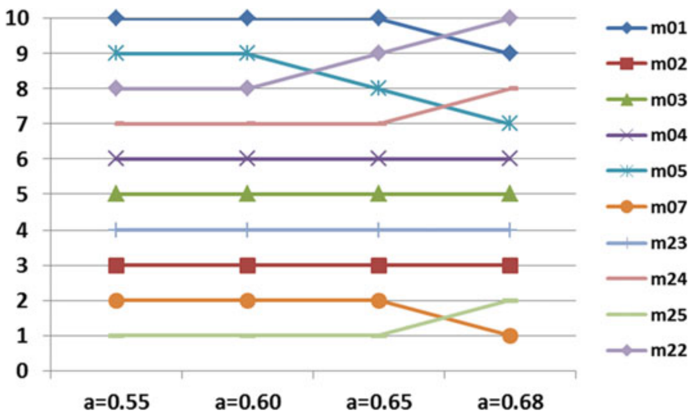


Fig. 2 Ranking positions of the top 10 MHE

Further research may investigate the outcomes of the proposed two-level Induced OWA procedure when considering the rational behaviour of the DM [46] or interval cross-efficiency [47].

Appendix

See Tables 14, 15, 16, 17, 18 and 19.

Table 14 Data set of 12 manufacturing systems [43]

DMU	Outputs				Inputs	
	y_1	y_2	y_3	y_4	x_1	x_2
S_1	42	45.3	14.2	30.1	17.02	5
S_2	39	40.1	13	29.8	16.46	4.5
S_3	26	39.6	13.8	24.5	11.76	6
S_4	22	36	11.3	25	10.52	4
S_5	21	34.2	12	20.4	9.5	3.8
S_6	10	20.1	5	16.5	4.79	5.4
S_7	14	26.5	7	19.7	6.21	6.2
S_8	25	35.9	9	24.7	11.12	6
S_9	4	17.4	0.1	18.1	3.67	8
S_{10}	16	34.3	6.5	20.6	8.93	7
S_{11}	43	45.6	14	31.1	17.74	7.1
S_{12}	27	38.7	13.8	25.4	14.85	6.2

Table 15 Data set of 14 bank branches [44]

DMU	Outputs				Inputs		
	y_1	y_2	y_3	y_4	x_1	x_2	x_3
B_1	484,000	4,139,100	59,860	2,951,430	140,000	42,900	87,500
B_2	384,000	1,685,500	139,780	3,336,860	48,800	17,400	37,900
B_3	209,000	1,058,900	65,720	3,570,050	36,600	14,200	29,800
B_4	157,000	879,400	27,340	2,081,350	47,100	9300	26,800
B_5	46,000	370,900	18,920	1,069,100	32,600	4600	19,600
B_6	272,000	667,400	34,750	2,660,040	50,800	8300	18,900
B_7	53,000	465,700	20,240	1,800,250	40,800	7500	20,400
B_8	250,000	642,700	43,280	2,296,740	31,900	9200	21,400
B_9	407,000	647,700	32,360	1,981,930	36,400	76,000	21,000
B_{10}	72,000	402,500	19,930	2,284,910	25,700	7900	19,000
B_{11}	105,000	482,400	49,320	2,245,160	44,500	8700	21,700
B_{12}	94,000	511,000	26,950	2,303,000	42,300	8900	25,800
B_{13}	84,000	287,400	34,940	1,141,750	40,600	5500	19,400
B_{14}	199,000	694,600	67,160	3,338,390	76,100	11,900	32,800

Table 16 Data set of 14 airlines [45]

DMU	Outputs		Inputs		
	y_1	y_2	x_1	x_2	x_3
A_1	26,677	697	5723	3239	2003
A_2	3081	539	5895	4225	4557
A_3	124,055	1266	24,099	9560	6267
A_4	64,734	1563	13,565	7499	3213
A_5	23,604	513	5183	1880	783
A_6	95,011	572	19,080	8032	3272
A_7	22,112	969	4603	3457	2360
A_8	52,363	2001	12,097	6779	6474
A_9	26,504	1297	6587	3341	3581
A_{10}	19,277	972	5654	1878	1916
A_{11}	41,925	3398	12,559	8098	3310
A_{12}	27,754	982	5728	2481	2254
A_{13}	31,332	543	4715	1792	2485
A_{14}	122,528	1404	22,793	9874	4145

Table 17 Cross-efficiency matrix ***E*** for MHEs

MHE	m_{01}	m_{02}	m_{03}	m_{04}	m_{05}	m_{06}	m_{07}	m_{08}	m_{09}	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}	m_{17}	m_{18}	m_{19}	m_{20}	m_{21}	m_{22}	m_{23}	m_{24}	m_{25}
m_{01}	0.808	0.941	0.895	1	0.934	0.791	1	0.271	0.268	0.270	0.227	0.325	0.377	0.365	0.360	0.312	0.048	0.038	0.041	0.047	0.049	0.527	1	0.568	0.769
m_{02}	0.757	1	0.930	0.829	0.702	0.732	1	0.607	0.686	0.662	0.521	0.533	0.519	0.452	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.967	0.798	1
m_{03}	0.757	1	0.930	0.829	0.702	0.732	1	0.557	0.616	0.599	0.476	0.508	0.502	0.436	0.565	0.504	0.185	0.158	0.174	0.193	0.206	0.754	0.967	0.783	1
m_{04}	0.808	0.941	0.895	1	1	0.791	1	0.271	0.268	0.270	0.227	0.325	0.379	0.384	0.360	0.312	0.048	0.038	0.041	0.047	0.049	0.527	1	0.568	0.769
m_{05}	0.792	0.920	0.881	1	1	0.779	1	0.241	0.230	0.237	0.200	0.314	0.394	0.405	0.346	0.300	0.038	0.029	0.032	0.037	0.039	0.445	0.904	0.483	0.661
m_{06}	0.808	0.941	0.895	1	0.934	0.791	1	0.271	0.268	0.270	0.227	0.325	0.377	0.365	0.360	0.312	0.048	0.038	0.041	0.047	0.049	0.527	1	0.568	0.769
m_{07}	0.808	1	0.930	1	1	0.791	1	0.571	0.630	0.614	0.488	0.523	0.519	0.516	0.581	0.520	0.185	0.158	0.174	0.193	0.206	0.754	1	0.783	1
m_{08}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{09}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{10}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{11}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{12}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{13}	0.743	1	0.929	0.806	0.688	0.719	1	0.571	0.630	0.614	0.488	0.523	0.519	0.452	0.581	0.520	0.185	0.157	0.172	0.192	0.205	0.643	0.710	0.669	0.835
m_{14}	0.702	0.922	0.878	0.826	0.878	0.691	1	0.310	0.294	0.306	0.255	0.398	0.503	0.516	0.434	0.382	0.048	0.036	0.040	0.045	0.049	0.193	0.220	0.210	0.260
m_{15}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{16}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{17}	0.412	0.557	0.514	0.437	0.357	0.396	0.546	0.393	0.478	0.446	0.343	0.310	0.286	0.239	0.346	0.311	0.342	0.460	0.516	0.493	0.516	0.821	0.857	0.788	1
m_{18}	0.412	0.557	0.514	0.437	0.357	0.396	0.546	0.393	0.478	0.446	0.343	0.310	0.286	0.239	0.346	0.311	0.342	0.460	0.516	0.493	0.516	0.821	0.857	0.788	1
m_{19}	0.412	0.557	0.514	0.437	0.357	0.396	0.546	0.393	0.478	0.446	0.343	0.310	0.286	0.239	0.346	0.311	0.342	0.460	0.516	0.493	0.516	0.821	0.857	0.788	1
m_{20}	0.412	0.557	0.514	0.437	0.357	0.396	0.546	0.393	0.478	0.446	0.343	0.310	0.286	0.239	0.346	0.311	0.342	0.460	0.516	0.493	0.516	0.821	0.857	0.788	1
m_{21}	0.412	0.557	0.514	0.437	0.357	0.396	0.546	0.393	0.478	0.446	0.343	0.310	0.286	0.239	0.346	0.311	0.342	0.460	0.516	0.493	0.516	0.821	0.857	0.788	1

(continued)

Table 17 (continued)

MHE	m_{01}	m_{02}	m_{03}	m_{04}	m_{05}	m_{06}	m_{07}	m_{08}	m_{09}	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}	m_{17}	m_{18}	m_{19}	m_{20}	m_{21}	m_{22}	m_{23}	m_{24}	m_{25}
m_{22}	0.412	0.557	0.514	0.437	0.357	0.396	0.546	0.393	0.478	0.446	0.343	0.310	0.286	0.239	0.346	0.311	0.342	0.460	0.516	0.493	0.516	0.821	0.857	0.788	1
m_{23}	0.808	0.998	0.928	1	0.991	0.791	1	0.540	0.594	0.579	0.461	0.498	0.495	0.431	0.554	0.494	0.271	0.366	0.410	0.391	0.409	0.800	1	0.778	1
m_{24}	0.742	1	0.927	0.799	0.673	0.717	0.994	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.236	0.209	0.230	0.252	0.270	0.778	0.851	0.798	1
m_{25}	0.760	1	0.930	0.838	0.710	0.735	1	0.607	0.686	0.662	0.521	0.533	0.517	0.445	0.593	0.531	0.342	0.460	0.516	0.493	0.516	0.821	1	0.798	1

Table 18 Preference voting matrix Θ for MHE

MHE	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
m_{01}	13	1	0	0	1	0	0	1	1	3	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0
m_{02}	9	0	4	0	1	0	0	0	0	1	1	4	0	1	1	1	0	1	1	0	0	0	0	0	0
m_{03}	9	0	0	3	0	1	0	0	0	0	2	1	1	2	0	1	2	2	1	0	0	0	0	0	0
m_{04}	9	4	0	0	0	0	0	0	2	0	1	0	0	2	1	1	2	2	1	0	0	0	0	0	0
m_{05}	9	0	0	1	2	0	1	0	1	0	2	1	0	0	0	1	2	2	1	2	0	0	0	0	0
m_{06}	12	1	0	2	0	0	0	1	0	1	1	0	2	0	1	0	3	0	0	0	1	0	0	0	0
m_{07}	9	1	3	0	2	1	0	0	1	0	2	0	0	1	0	0	2	2	1	0	0	0	0	0	0
m_{08}	9	0	0	0	0	1	1	0	0	1	1	0	1	4	1	2	2	1	0	0	1	0	0	0	0
m_{09}	9	3	0	0	0	0	0	4	2	0	1	0	1	0	0	0	2	1	0	0	2	0	0	0	0
m_{10}	9	0	0	0	1	1	1	3	0	0	1	1	0	0	0	1	1	2	1	0	2	1	0	0	0
m_{11}	11	1	1	1	0	1	1	0	0	0	1	1	0	1	1	1	0	0	2	0	2	0	0	0	0
m_{12}	9	0	1	3	2	0	1	0	1	0	1	0	0	0	0	2	0	1	0	2	2	0	0	0	0
m_{13}	9	0	0	1	0	1	1	0	1	1	1	0	1	0	0	1	3	2	1	1	1	0	0	0	0
m_{14}	9	2	1	1	0	1	0	0	0	1	2	1	0	2	1	1	1	1	0	0	1	0	0	0	0
m_{15}	9	0	0	1	2	0	1	0	3	1	2	1	0	2	1	0	1	1	0	0	0	0	0	0	0
m_{16}	13	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1	0	0	1	0	1	0	0	0
m_{17}	9	0	3	0	1	1	0	1	0	2	1	0	1	0	1	2	1	0	0	1	0	1	0	0	0
m_{18}	9	0	0	0	0	1	1	0	0	1	1	1	1	2	1	0	1	2	0	0	2	1	1	0	0
m_{19}	9	3	0	0	0	0	0	0	0	0	0	2	1	1	1	2	2	2	2	0	0	0	0	0	0
m_{20}	9	0	0	0	1	0	1	2	0	0	3	1	0	1	0	1	1	1	1	0	1	0	1	1	0
m_{21}	12	0	1	2	0	1	2	0	1	0	0	0	3	0	1	1	0	0	0	0	0	0	0	0	1

(continued)

Table 18 (continued)

MHE	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
m_{22}	9	0	2	0	0	2	2	1	0	0	2	0	0	1	0	1	2	0	2	0	0	0	0	0	1
m_{23}	9	0	0	0	0	3	1	0	0	1	1	0	0	1	1	1	3	0	0	1	0	0	1	2	0
m_{24}	9	0	0	0	2	0	0	0	4	0	2	0	0	1	0	1	2	0	2	1	0	0	1	0	0
m_{25}	9	0	0	0	1	0	2	2	0	1	1	0	1	2	0	1	2	0	1	1	0	0	0	0	1

Table 19 OWA weights generated for different subjectivity levels α with K=25

OWA	α = 0.55	α = 0.60	α = 0.65	α = 0.68
γ_1	0.05108	0.06215	0.07323	0.07988
γ_2	0.05015	0.06031	0.07046	0.07655
γ_3	0.04923	0.05846	0.06769	0.07323
γ_4	0.04831	0.05662	0.06492	0.06991
γ_5	0.04738	0.05477	0.06215	0.06658
γ_6	0.04646	0.05292	0.05938	0.06326
γ_7	0.04554	0.05108	0.05662	0.05994
γ_8	0.04462	0.04923	0.05385	0.05662
γ_9	0.04369	0.04738	0.05108	0.05329
γ_{10}	0.04277	0.04554	0.04831	0.04997
γ_{11}	0.04185	0.04369	0.04554	0.04665
γ_{12}	0.04092	0.04185	0.04277	0.04332
γ_{13}	0.04000	0.04000	0.04000	0.04000
γ_{14}	0.03908	0.03815	0.03723	0.03668
γ_5	0.03815	0.03631	0.03446	0.03335
γ_{16}	0.03723	0.03446	0.03169	0.03003
γ_{17}	0.03631	0.03262	0.02892	0.02671
γ_{18}	0.03538	0.03077	0.02615	0.02338
γ_{19}	0.03446	0.02892	0.02338	0.02006
γ_{20}	0.03354	0.02708	0.02062	0.01674
γ_{21}	0.03262	0.02523	0.01785	0.01342
γ_{22}	0.03169	0.02338	0.01508	0.01009
γ_{23}	0.03077	0.02154	0.01231	0.00677
γ_{24}	0.02985	0.01969	0.00954	0.00345
γ_{25}	0.02892	0.01785	0.00677	0.00012

References

1. Sexton, T.R., Silkman, R.H., Hogan, A.J.: Data envelopment analysis: critique and extensions. In: Silkman, R.H. (ed.) Measuring Efficiency: An Assessment of Data Envelopment Analysis. Jossey Bass, San Francisco, CA (1986)
2. Wu, J., Chu, J., Sun, J., Zhu, Q.: DEA cross-efficiency evaluation based on Pareto improvement. Eur. J. Oper. Res. **248**(2), 571–579 (2016)
3. Oukil, A., Al-Zidi, A.: Benchmarking the hotel industry in Oman through a three-stage DEA based procedure. J. Arts Soc. Sci. **9**(2), 5–23 (2018)
4. Chen, H.: Average lexicographic efficiency for data envelopment analysis. Omega **74**, 82–91 (2018)
5. Al-Mezeini, N. Oukil, A., Al-Ismaili, A.: Investigating the efficiency of greenhouse production in Oman: a two-stage approach based on Data Envelopment Analysis and double bootstrapping. J. Cleaner Prod. **247**, 119160 (2021). In press. https://doi.org/10.1016/j.jclepro.2019.119160

6. Oral, M., Oukil, A., Malouin, J.-L., Kettani, O.: The appreciative democratic voice of DEA: a case of faculty academic performance evaluation. Socio Econ. Plann. Sci. **48**, 20–28 (2014)
7. Mahdiloo, M., Saen, R.F., Lee, K.H.: Technical, environmental and eco-efficiency measurement for supplier selection: an extension and application of data envelopment analysis. Int. J. Prod. Econ. **168**, 279–289 (2015)
8. Oukil A, Govindaluri S.M.: A systematic approach for ranking football players within an integrated DEA-OWA framework. Manag. Decis. Econ. **38**, 1125–1136 (2017)
9. Oukil, A., Amin, G.R.: Maximum appreciative cross-efficiency in DEA: a new ranking method. Comput. Ind. Eng. **81**, 14–21 (2015)
10. Banihashemi, S., Sanei, M.: Cross efficiency evaluation with negative data in selecting the best of portfolio using OWA operator weights. Int. J. Data Envelopment Anal. **3**, 633–651 (2015)
11. Dotoli, M., Epicoco, N., Falagario, M., Sciancalepore, F.: A stochastic cross-efficiency data envelopment analysis approach for supplier selection under uncertainty. Int. Trans. Oper. Res. **23**, 725–748 (2016)
12. Amin, G.R., Oukil, A.: Gangless cross-evaluation in DEA: an application to stock selection. RAIRO-Oper. Res. **53**(2), 645–655 (2019)
13. Oukil, A., El-Bouri, A.: Ranking dispatching rules in multi-objective dynamic flow shop scheduling: a multi-faceted perspective. Int. J. Prod. Res. **59**(2), 388–411 (2021)
14. Hassan, M., Oukil, A.: Design of efficient systems of commercial material handling equipment for supply chain and logistics facilities using DEA. Int. J. Logistics Syst. Manage. **39**(2), 241–272 (2021)
15. Wu, J., Liang, L., Zha, Y.: Determination of the weights of ultimate cross efficiency based on the solution of nucleolus in cooperative game. Syst. Eng. Theor. Pract. **28**, 92–97 (2008)
16. Wu, J., Liang, L., Yang, F.: Determination of the weights for the ultimate cross efficiency using Shapley value in cooperative game. Expert Syst. Appl. **36**, 872–876 (2009)
17. Song, L., Liu, F.: An improvement in DEA cross-efficiency aggregation based on the Shannon entropy. Int. Trans. Oper. Res. **25**, 705–714 (2018)
18. Yang, G.L., Yang, J.B., Liu, W.B., Li, X.X.: Cross-efficiency aggregation in DEA models using the evidential-reasoning approach. Eur. J. Oper. Res. **231**(2), 393–404 (2013)
19. Wang, Y.M., Chin, K.S.: The use of OWA operator weights for cross-efficiency aggregation. Omega **39**, 493–503 (2011)
20. León, T., Ramón, N., Ruiz, J.L., Sirvent, I.: Using Induced Ordered Weighted Averaging IOWA operators for aggregation in cross-efficiency evaluations. Int. J. Intell. Syst. **29**, 1100–1116 (2014)
21. Angiz, M.Z., Mustafa, A., Kamali, M.J.: Cross-ranking of decision making units in data envelopment analysis. Appl. Math. Model. **37**(1–2), 398–405 (2013)
22. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. Eur. J. Oper. Res. **2**(6), 429–444 (1978)
23. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage. Sci. **30**(9), 1078–1092 (1984)
24. Cooper, W.W., Seiford, L.M., Tone, H.: Data envelopment analysis: A comprehensive text with models, applications, references, and DEA-Solver software. Kluwer Academic Publisher, New York (2002)
25. Emrouznejad, A.: Advances in data envelopment analysis. Ann. Oper. Res. **214**(1), 1–4 (2014)
26. Emrouznejad, A., Yang, G.L.: A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. Socioecon. Plann. Sci. **61**, 4–8 (2018)
27. Oukil, A., Soltani, A.A., Boutaghane, H., et al.: A Surrogate Water Quality Index to assess groundwater using a unified DEA-OWA framework. Environ. Sci. Pollut. Res. (2021). https://doi.org/10.1007/s11356-021-13758-0
28. Soltani, A.A., Oukil, A., Boutaghane, H., Bermad, A., Boulassel, M.R.: A new methodology for assessing water quality, based on data envelopment analysis: application to Algerian dams. Ecol. Ind. **121**, 106952 (2021). https://doi.org/10.1016/j.ecolind.2020.106952
29. Sow, O., Oukil, A., Ndiaye, B.M., Marcos, A.: Efficiency analysis of public transportation subunits using DEA and bootstrap approaches–Dakar Dem Dikk case study. J. Math. Res. **8** (2016)

30. Oukil, A.: Ranking via composite weighting schemes under a DEA cross-evaluation framework. Comput. Ind. Eng. **117**, 217–224 (2018)
31. Oral, M., Amin, G.R., Oukil, A.: Cross-efficiency in DEA: A maximum resonated appreciative model. Measurement **63**, 159–167 (2015)
32. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Trans. Syst. Man Cybern. **18**, 183–190 (1988)
33. Yager, R.R.: An approach to ordinal decision making. Int. J. Approximate Reasoning **12**(3–4), 237–261 (1995)
34. Saeidi, R.G., Oukil, A., Amin, G.R., Raissi, S.: Prioritization of textile fabric defects using ordered weighted averaging operator. Int. J. Adv. Manuf. Technol. **76**(5–8), 745–752 (2015)
35. Amin, G.R., El-Bouri, A.: A combined OWA-DEA method for dispatching rule selection. Comput. Ind. Eng. **88**, 470–478 (2015)
36. Oukil, A.: Embedding OWA under preference ranking for DEA cross-efficiency aggregation: issues and procedures. Int. J. Intell. Syst. **34**, 947–965 (2019)
37. Oukil, A.: Designing efficient material handling systems: a two-stage approach based on DEA cross-efficiency. In: 5th International Conference on Logistics Operations Management (GOL), pp. 1–5 (2020). https://doi.org/10.1109/GOL49479.2020.9314754
38. Emrouznejad, A., Marra, M.: Ordered weighted averaging operators 1988–2014: a citation-based literature survey. Int. J. Intell. Syst. **29**, 994–1014 (2014)
39. Oukil, A.: Exploiting value system multiplicity and preference voting for robust ranking. Omega, **94**, 102048 (2020). https://doi.org/10.1016/j.omega.2019.03.006
40. Yager, R.R., Filev, D.P.: Induced ordered weighted averaging operators. EEE Trans. Syst. Man Cybern. B **29**(2), 141–150 (1999)
41. Aggarwal, M.: A new family of induced OWA operators. Int. J. Intell. Syst. **30**(2), 170–205 (2015)
42. Jin, L., Yager, R.R., Špirková, J., Mesiar, R., Paternain, D., Bustince, H.: OWA aggregation with dual preferences for basic uncertain information. J. Intell. Fuzzy Syst. (Preprint), 1–10 (2021)
43. Shang, J., Sueyoshi, T.: A unified framework for the selection of a flexible manufacturing system. Eur. J. Oper. Res. **85**(2), 297–315 (1995)
44. Sherman, H.D., Gold, F.: Bank branch operating efficiency: evaluation with DEA. J. Bank. Finance **9**, 297–315 (1985)
45. Tofallis, C.: Input efficiency profiling: an application to airlines. Comput. Oper. Res. **24**(3), 253–258 (1997)
46. Chen, L., Wang, Y.M., Huang, Y.: Cross-efficiency aggregation method based on prospect consensus process. Ann. Oper. Res. **288**(1), 115–135 (2020)
47. Zhang, X., Xia, Q., Yang, F., Song, S., Ang, S.: Interval cross-efficiency for ranking decision making units using the stochastic multicriteria acceptability analysis-evidential reasoning approach. Comput. Ind. Eng. **156**, 107222 (2021)

Simulation Optimization for Multi-product (s, S) Inventory Policy with Stochastic Demand

Ilkay Saracoglu

Abstract In this study, the inventory management problem of a wholesaler company that supplies and distributes products according to uncertain demand is discussed. This study constructs a multi-product multi-period (s, S) inventory policy considering items with uncertain demand and determines the maximum inventory level (S) and reorder point (s) for each item to minimize the total inventory cost for the wholesaler company in question, considering budget constraint. (s, S) policy means that an order is triggered as soon as the inventory position declines to or below the reorder point (s) and difficult to determine the (S) values. The order size is chosen so that the inventory position increases to (S). Since it is a multi-product and multi-period system, and demand has stochastic structure the results getting from mathematical model are validated by using simulation optimization model.

Keywords Multi-period ınventory policy · Continuous review policy · Stochastic demand · Simualtion optimization

1 Introduction

Inventory management problems have a significant place in literature since they play an important role in supply chain management. The principal objective of inventory management is to improve the level of customer service and to reduce the costs related to inventory. It should be noted that inventory models always attempt to attain an answer to the question of when and how much to order. Thus, inventory management addresses the question of how much inventory should be kept in stock so that the desired service level could be achieved. Supply chain management, on the other hand, connects manufacturing and storage activities by way of transportation and balances demand, supply, and cash flow. Inventory management determines the necessary inventory level to counter balance the cost of capital in the case of keeping excessive inventory with the penalty cost occurring with inventory shortage. Dealing

I. Saracoglu (✉)
Department of Industrial Engineering, Haliç University, Eyupsultan, Istanbul, Turkey
e-mail: ilkays@sbbdanismanlik.com

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_48

with conflicting objectives, effective inventory control could provide a significant advantage. The key factor affecting the solution is the demand structure, whether it is deterministic or probabilistic. In real life, demand usually has a probabilistic structure. However, in some cases, the simpler deterministic approach might be adopted to come to a solution. Another important factor for inventory models is cost. A further influential factor is lead time. This can be a random or fixed period of time. As well as these factors, it might be indispensable to supplement the model with certain constraint conditions. One of the most important conditions is the budget. There might also be a constraint condition with respect to storage area. Some products might deteriorate in time. Hence, shelf life is a significant element, as well.

This paper develops a solution methodology for multi-product multi-period (*s, S*) inventory models that calculates the optimal maximum inventory level and reorder point under the constraints of budget with stochastic demand structure. The approach used to solve this problem consists of two stages. In the first stage, a mathematical model is presented for the inventory policy with continuous review (s, S) inventory system considering that the products have a deterministic demand structure. As a result of solving this model with LINGO optimization software, optimum (s, S) level information is obtained for each product. In the second stage, the ARENA simulation model and simulation optimization approach with stochastic demand structure, which is not taken into account in the current mathematical model, are presented to test whether the minimum cost is reached under the desired budget constraint, using the information obtained from the first stage.

The remainder of this paper is organized as follows: Literature review is presented in Sect. 2. Problem statement and formulation are developed in Sect. 3. An illustrative experimental study is given in Sect. 4. Finally, conclusions and recommendations for future research are provided in Sect. 5.

2 Literature Review

In this paper, a multi-product multi-period (s, S) policy to minimize average total inventory cost of the wholesaler company is proposed by considering stochastic demand and budget constraint. This has been one of the most commonly applied policies in the literature. To begin with, [1] analyzed continuous review (s, S) policy where demand is defined as a discrete valued process. He presumed that the (s, S) policy is allowed for complete backlog of unsatisfied demand. He considered that the procurement cost consists of a fixed cost, a variable cost and holding cost. Next, [2] analyzed the continuous review inventory system with constant lead time and complete backlog of unsatisfied demand under (s, S) policy. On the other hand, considered continuous review perishable inventory models with lost-sales and backlogging [3]. Expected average cost is tried to be minimized under the (s, S) inventory policy with zero lead time and fixed lifetime assumptions and the demand shows a stationary Poisson process. The (s, S) policies applied on perishable items have been studied by researchers such as [1, 3–6].

When multi-period studies are reviewed, they are observed to be mostly used for lot-sizing problems in production. Firstly, [7] addressed (Q, r) policy assuming the multi-period inventory problem with single item. They tried to minimize the cost without any constraints. Veinott [8] applied a dynamic nonstationary multi-product inventory model and considered backlogged. Lapierre and Ruiz [9] analyzed the logistic system of a hospital to minimize the sum of inventory and labor costs with multi-product. Saracoglu et al. [10] developed multi-product multi-period (Q, r) inventory model with budget, storage and lifetime constraints. Liu [11] analyzed an economic lot sizing (ELS) problem for single item and multiple periods with both upper and lower inventory bounds. Qiu et al. [12] proposed robust dynamic model for a single product with considering uncertainty demand according to the (s, S) policy. Movahed and Zhang [13] developed optimal (s, S) inventory policy for three-level supply chain with uncertain demand and lead time. Kleijnen and Wan [14] illustrated simulation optimization for (s, S) inventory policy to minimize the inventory cost excluding shortage cost with constraint of service level. Noordhoek et al. [15] proposed scatter search based simulation–optimization model to determine the (s, S) inventory polices for multi-echelon distribution networks by considering fill rate constraint.

3 Problem Statement and Formulation

3.1 Proposed Mathematical Method

This study solves inventory problem of a wholesaler by using multi-period (s, S) model because of the uncertain demand structure. In a way similar to (Q, r) policy, (s, S) inventory policy is a system that procures order when the inventory level drops to the level of reorder point. Yet, unlike (Q, r) policy, order amount (quantity) is hereby calculated by the equation $Q = S - s$. S, is the order-up-to level. In this paper, we assume that multiple products, lead time is fixed and constant, shortages are allowed. The problem is to formulate a model that minimizes the average total inventory cost (ATIC) subject to constraints on budget. The following notations are used to illustrate the ILP formulation of the proposed multi-product multi-period (*s, S*) continuous review inventory model.

Sets and Indices:

I number of products, where, $i = 1, 2, \ldots, I$.
T number of planning periods, where, $t = 1, 2, \ldots, T$.

Parameters:

$d_{i,t}$ demand for product *i* in period *t*.
md_i expected mean demand during the lead time for product *i*.
L_i lead time for product *i*.

pc_i	unit purchasing price for product *i*.
oc_i	ordering cost for product *i,* per order.
shc_i	unit shortage cost for product *i*.
hc_i	unit holding cost for product *i,* per period.
$I_{i,0}$	initial inventory level for product *i* in period 0.
ub_s_i	the upper bound of the reorder point for product *i*.
ub_S_i	the upper bound of the maximum inventory level for product *i*.
$Budget$	available budget limit.
M	big number.

Decision variables:

$I_{i,t}$	ending inventory level for product *i* in period t.
$I^+_{i,t}$	positive inventory level for product *i* in period *t*, $I^+_{i,t} = max\{0, I_{i,t}\}$.
$I^-_{i,t}$	inventory shortage for product *i* in period *t*, $I^-_{i,t} = max\{0, -I_{i,t}\}$.
$Q_{i,t}$	order quantity for product *i* in period t.
S_i	maximum inventory level for product *i*.
s_i	reorder point for product *i*.
$X_{i,t}$	1, if beginning inventory level of product *i* is less than reorder point in period t, 0 otherwise.
$Z_{i,t}$	1, if ending inventory level of product *i* in period *t* is positive, 0 otherwise.

Our aim is to minimize ATIC; therefore, the objective function is the total of the holding cost (THC), ordering cost (TOC) and shortage cost (TSC) for one period of the planning horizon. Consequently, the objective function (1) of the multi-product, multi-period (*s, S*) policy model is given in the equation below:

$$MinZ/T = \sum_{i=1}^{I}\sum_{t=1}^{T} hc_i \cdot I^+_{i,t} + \sum_{i=1}^{I}\sum_{t=1}^{T} U_{i,t} \cdot os_i + \sum_{i=1}^{I}\sum_{t=1}^{T} shc_i \cdot I^-_{i,t} \quad (1)$$

Constraints
If

$$I_{i,t-1} \le s_i \quad i = 1, \ldots, I, t = 1, \ldots, T \quad (2)$$

Then

$$Q_{i,t} - I^+_{i,t} + I^-_{i,t} + I^+_{i,t-1} - d_{i,t} = 0 \quad i = 1, \ldots, I, t = 1, \ldots, T \quad (3)$$

Else

$$I^+_{i,t-1} - I^+_{i,t} + I^-_{i,t} - d_{i,t} = 0 \quad i = 1, \ldots, I, t = 1, \ldots, T \quad (4)$$

and

$$Q_{i,t} = 0 \quad i = 1, \ldots, I, t = 1, \ldots, T \quad (5)$$

$$I^{+}_{i,t-1} - s_i + 1 \leq M \cdot (1 - X_{i,t}) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{6}$$

$$s_i - I^{+}_{i,t-1} \leq M \cdot X_{i,t} \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{7}$$

$$Q_{i,t} - I^{+}_{i,t} + I^{-}_{i,t} + I^{+}_{i,t-1} - d_{i,t} \geq -M \cdot (1 - X_{i,t}) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{8}$$

$$Q_{i,t} - I^{+}_{i,t} + I^{-}_{i,t} + I^{+}_{i,t-1} - d_{i,t} \leq M \cdot \left(1 - X_{i,t}\right) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{9}$$

$$I^{+}_{i,t-1} - I^{+}_{i,t} + I^{-}_{i,t} - d_{i,t} \geq -M \cdot \left(1 - X_{i,t}\right) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{10}$$

$$I^{+}_{i,t-1} - I^{+}_{i,t} + I^{-}_{i,t} - d_{i,t} \leq M \cdot \left(1 - X_{i,t}\right) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{11}$$

$$Q_{i,t} \leq M \cdot X_{i,t} \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{12}$$

$$S_i - I^{+}_{i,t-1} - Q_{i,t} \geq -M \cdot (1 - X_{i,t}) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{13}$$

$$S_i - I^{+}_{i,t-1} - Q_{i,t} \leq M \cdot (1 - X_{i,t}) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{14}$$

$$I^{-}_{i,t} \leq M \cdot \mathrm{Z}(i, t) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{15}$$

$$I^{+}_{i,t} \leq M \cdot (1 - \mathrm{Z}(i, t)) \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{16}$$

$$I_{i,t} = I^{+}_{i,t} - I^{-}_{i,t} \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{17}$$

$$1/T \sum_{i=1}^{I} \sum_{t=1}^{T} Q_{i,t} \cdot pc_i \leq Budget \tag{18}$$

$$S_i \leq ub\_S_i \quad i = 1, \ldots, I \tag{19}$$

$$s_i \leq ub\_s_i \quad i = 1, \ldots, I \tag{20}$$

$$s_i - S_i \leq 0 \quad i = 1, \ldots, I \tag{21}$$

$$X_{i,t}, Z_{i,t} \in \{0, 1\} \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{22}$$

$$s_i, S_i, Q_{i,t}, I_{i,t}, I^{+}_{i,t}, I^{-}_{i,t} \geq 0, integer \quad i = 1, \ldots, I, t = 1, \ldots, T \tag{23}$$

Equation (2) provide that if the beginning inventory level is lower than the reorder point level, an order quantity $Q_{i,t}$ is opened. Equation (3) shows that inventory balance equation. Equation (4) ensures that if the beginning inventory level is greater than the reorder point, no order is placed. Equations (6–12) were transformed the model from nonlinear to linear structure. Equation (9) is used to determine whether an order has been placed or not. Equations (13) and (14) added to determine the order quantities. Equations (15–17) explain on-hand and shortage inventory level. Equation (18) defines the budget constraints. Equations (19) and (20) explain that the ordered quantities and reorder points should be less than or equal to the upper bound values specified for each product. Equation (21) ensures that the reorder point is less than the maximum inventory level. Equation (22) define the binary variables. Equation (23) ensures that the integer restrictions for all the variables are used in the model.

3.2 Proposed Simulation Optimization Method

As the number of products and the number of periods increase, it is not possible to obtain optimum results in acceptable times with the mathematical model. Creating a stochastic mathematical model is quite complex. Therefore, the mathematical (ILP) model has been developed considering the deterministic demand structure. By using the developed ILP, the maximum inventory level and reorder point levels were obtained for each product. The optimum results obtained were used as input values in the simulation. In the simulation, the demand structure is considered as stochastic. In the simulation, it was possible to work in a shorter time and with more periods. The model established in the simulation is given in Fig. 1.

The simulation model was developed in Arena Version 16.00 like [16]'s book (s, S), the inventory policy model gives an example for a single product. Using this example, a model was developed for multiple products. In addition, another difference of the model is that if the inventory level drops to a negative level at the end of the period, the next inventory at the beginning of the period is assigned as 0. Thus, the backorder is not allowed, and the unmet demand is lost. In this section, simulation approach, which has an embedded simulation optimization by using the OptQuest for ARENA for deciding the optimal or near optimal (*s, S*) inventory policy. OptQuest engine combines meta-heuristics of tabu search, scatter search, neural networks algorithms in one search algorithm. Simulation optimization was used for (s, S) inventory policy by [17–19].

4 Experimental Study

The company assumed in this paper is a wholesaler performing drug distribution with a single regional storage. We selected four products to show the performance of the

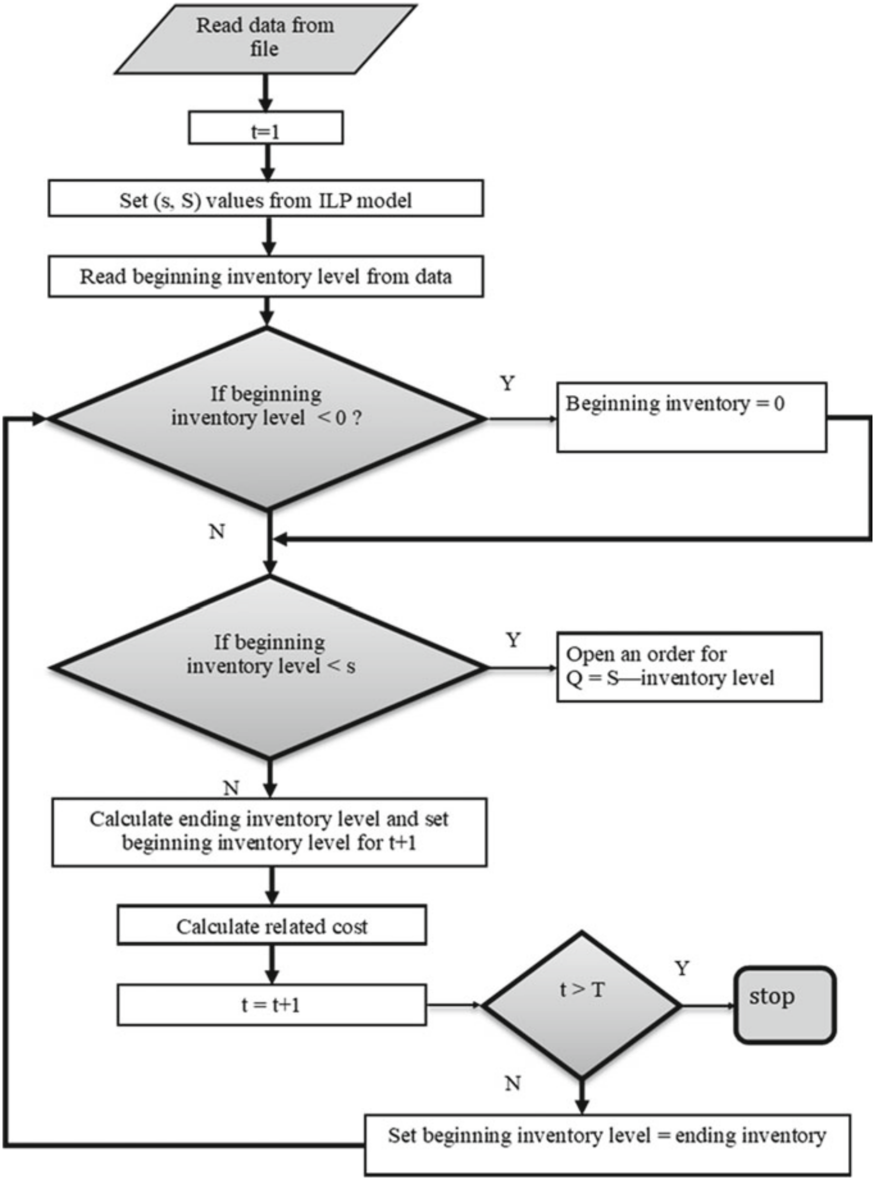


Fig. 1 Flowchart of the simulation model

model. Input data of the products are shown in Table 1. Firstly, ILP model are solved and obtained the solutions according to the deterministic demand for each product by using LINGO 18.0 software. If number of products and periods are increased, global optimal solution is not reached with a reasonable time. Because of the time restriction, 24 period was selected to solve the model. Next, we determined the lower and upper bound values for our (s_i, S_i) decision variables. The lower (lb_s_i, lb_S_i) and upper bound values (ub_s_i, ub_S_i) are listed in Table 2. Upper bound of s_i is determined according to the formulation of the expected demand during the lead time and safety stock by multiplying four times. Upper bound of S_i is calculated summation of the s_i and economic order quantity (EOQ) by multiplying four times. After calculating the (s_i, S_i) values according to the ILP model without considering the budget constraint, simulation model was constructed with stochastic demand data sets. The results of the ILP model is shown in Table 3. The simulation model solutions were obtained from an ARENA Student Version 16.0 software. All test problems were solved on an Intel ® Core™ i7-1051U CPU@ 1.80 GHz and 8.00 GB RAM. Normal distribution was considered for four products with mean demand and standard deviation of the lead time. These values are generated for all products (107, 46), (49, 19), (96, 45) and (161, 40), respectively.

Simulation model was run with 10 replications and 156 weeks. The average total inventory cost that the wholesaler will obtain at the end of the planning horizon is estimated as $96.72 under the $4000 budget constraints. This inventory cost level can be reached by placing orders of $S_i = \{414, 245, 398, 863\}$ order quantities at s_i

Table 1 Initial data for products

Products	1	2	3	4
L_i (week)	1	1	1	1
EOQ_i (unit)	269	167	251	677
pc_i ($)	12.83	15.30	13.17	3.04
hc_i ($)	0.09	0.11	0.09	0.02
oc_i ($)	30	30	30	30
shc_i ($)	14.14	16.83	14.49	3.34
$I_{i,0}$ (unit)	0	0	0	0
md_i (unit)	107	49	96	161
Standart deviation of lead time demand (sd_i)	46	19	45	40

Table 2 The lower and upper bounds for the (s, S) variables

Products	1	2	3	4
lb_S_i	0	0	0	0
ub_S_i	1543	890	1428	3404
lb_s_i	0	0	0	0
ub_s_i	469	224	423	694

Table 3 The results of the ILP model are illustrated in Table 2

Products	1	2	3	4	Total
s_i	118	30	97	48	
S_i	287	166	286	769	
Average purchasing cost	1353.53	612.64	1425.65	476.01	3867.83
Average total inventory cost	24.00	14.88	23.04	13.29	75.21

= {141, 75, 112, 241} inventory levels. Ten independent replications were applied for the proposed simulation optimization approach; the details of the best solution obtained are presented in Table 4. The computational time for an average replication was 2737.07 s. The result obtained by using OptQuest is shown in Fig. 2. This figure shows the graph of the simulation number and objective values for the $4000 budget constraints. Constructed simulation model in ARENA run with (s_i, S_i) values obtained from ILP model results for deterministic demand. Figure 3 shows that if we used ILP model solutions for stochastic demand, ATIC would be expected average $173.83. When the budget decreases, the amount of shortage increases. Whichever product has a low shortage cost, it tries to meet the budget constraint by not allowing it to purchase from that product. According to Table 4, if the budget is $2000, it is seen that no purchase was made because the product 4 was the lowest priced product. Average purchasing cost for each period is expected between the range of $3832 and $4297 with %95 significant level (Fig. 3).

Table 4 Solution results with different level of budget constraints

Budget constraints	Products	1	2	3	4	Total
Budget ≤ $4000	s	141	75	112	241	
	S	414	245	398	863	
	ATPC ($)	1385.32	806.39	1287.84	510.54	3990.11
	ATIC ($)	28.62	21.82	30.33	15.93	96.72
Budget ≤ $5000	s	148	79	135	245	
	S	424	216	461	876	
	ATPC ($)	1380.23	797.39	1276.90	518.00	3972.53
	ATIC ($)	28.77	21.62	29.38	15.87	95.65
Budget ≤ $3000	S	227	0	137	0	
	S	407	890	302	0	
	ATPC ($)	1383.03	0	1274.72	0	2657.76
	ATIC ($)	35.46	818.23	27.69	529.86	1411.26
Budget ≤ $2000	s	107	224	0	0	
	S	107	325	0	2179	
	ATPC ($)	1128.90	829.56	0	0	1958.46
	ATIC ($)	259.79	38.17	1333.82	529.86	2161.65

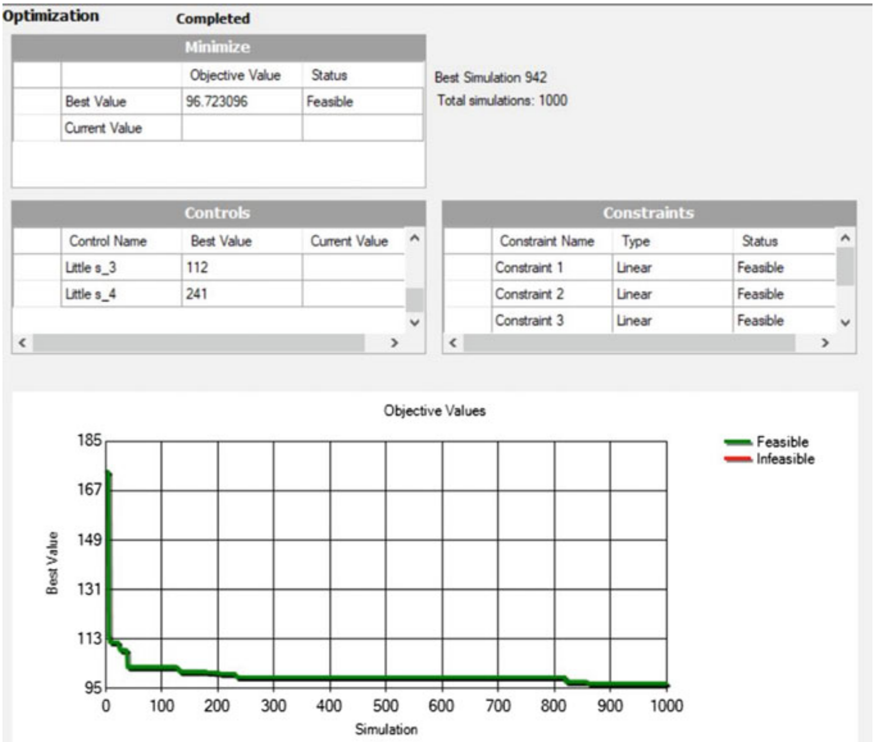


Fig. 2 The results from OptQuest

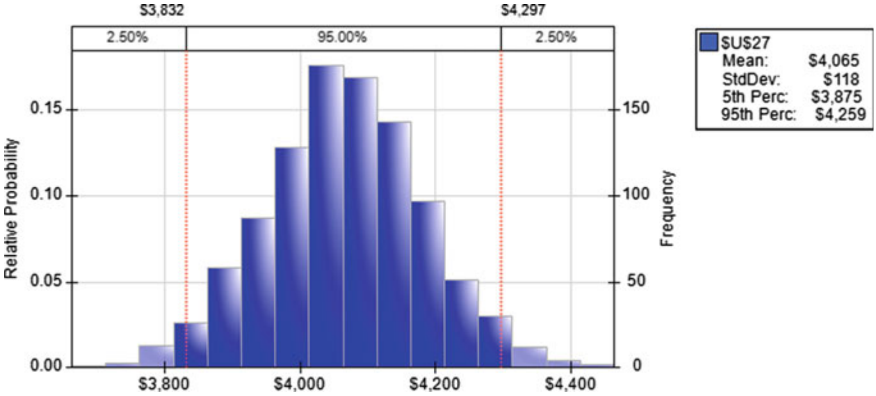


Fig. 3 Deviation of the total purchasing cost

5 Conclusion

In this paper, we formulated multi product and multi period (*s, S*) inventory models, with the objective of minimizing average total inventory cost under budget constraints. We found that an optimal solution cannot be reached with ILP in a reasonable computational time, if the number of products and periods increase. For this reason, we proposed a simulation optimization approach since the problem of the wholesaler case studied is a larger sized problem. We show that ILP model and simulation optimization how to use together to solve the inventory management problem very efficiently.

In future research, this model can further be generalized to include the variable lead time and consider different constraints such as capacity, lifetime, service level. In the proposed model, a solution can be sought by considering the safety stock level as the lower bound value of the reorder point. In addition, we will generalize the model to handle supply chain management problems including multi echelon in the model.

References

1. Sivazlian, B.D.: A continuous-review (s, S) inventory system with arbitrary interarrival distribution between unit demand. Oper. Res. **22**, 65–71 (1974)
2. Sahin, I.: On the stationary analysis of continuous review (s, S) inventory systems with constant lead time. Oper. Res. **27**, 717–729 (1979)
3. Weiss, H.J.: Optimal ordering policies for continuous review perishable inventory models. Oper. Res. **28**, 365–374 (1980). https://doi.org/10.1287/opre.28.2.365
4. Sivakumar, B.: A perishable inventory system with retrial demands and a finite population. J. Comput. Appl. Math. **224**, 29–38 (2009). https://doi.org/10.1016/j.cam.2008.03.041
5. Gürler, Ü., Özkaya, B.Y.: Analysis of the (s, S) policy for perishables with a random shelf life. IIE Trans. (Inst. Ind. Eng.) **40**, 759–781 (2008). https://doi.org/10.1080/07408170701730792
6. Lian, Z., Liu, X., Zhao, N.: A perishable inventory model with Markovian renewal demands. Int. J. Prod. Econ. **121**, 176–182 (2009). https://doi.org/10.1016/j.ijpe.2009.04.026
7. Veinott, A.F., Hadley, G., Whitin, T.M.: Analysis of Inventory Systems. Prentice-Hall, Englewood Cliffs, New York (1963)
8. Veinott, A.F.: Optimal policy for a multi-product, dynamic, nonstationary inventory problem. Manage. Sci. **12**, 206–222 (1965). https://doi.org/10.1287/mnsc.12.3.206
9. Lapierre, S.D., Ruiz, A.B.: Scheduling logistic activities to improve hospital supply systems. Comput. Oper. Res. **34**, 624–641 (2007). https://doi.org/10.1016/j.cor.2005.03.017
10. Saracoglu, I., Topaloglu, S., Keskinturk, T.: A genetic algorithm approach for multi-product multi-period continuous review inventory models. Expert Syst. Appl. **41**, 8189–8202 (2014). https://doi.org/10.1016/j.eswa.2014.07.003
11. Liu, T.: Economic lot sizing problem with inventory bounds. Eur. J. Oper. Res. **185**, 204–215 (2008). https://doi.org/10.1016/j.ejor.2006.12.032
12. Qiu, R., Sun, M., Lim, Y.F.: Optimizing (s, S) policies for multi-period inventory models with demand distribution uncertainty: robust dynamic programing approaches. Eur. J. Oper. Res. **261**, 880–892 (2017). https://doi.org/10.1016/j.ejor.2017.02.027
13. Movahed, K.K., Zhang, Z.H.: Robust design of (s, S) inventory policy parameters in supply chains with demand and lead time uncertainties. Int. J. Syst. Sci. **46**, 2258–2268 (2015). https://doi.org/10.1080/00207721.2013.860637

14. Kleijnen, J.P.C., Wan, J.: Optimization of simulated systems: OptQuest and alternatives. Simul. Model. Pract. Theory **15**, 354–362 (2007). https://doi.org/10.1016/j.simpat.2006.11.001
15. Noordhoek, M., Dullaert, W., Lai, D.S.W., de Leeuw, S.: A simulation–optimization approach for a service-constrained multi-echelon distribution network. Transp. Res. Part E Logist. Transp. Rev. **114**, 292–311 (2018). https://doi.org/10.1016/j.tre.2018.02.006
16. Kelton, W.D., Sadowski, R.P., Swets, N.B.: Simulation with Arena International Edition. McGraw-Hill (2010)
17. Ekren, B.Y., Ornek, M.A.: Determining optimum (s, S) levels of floor stock items in a paint production environment. Simul. Model. Pract. Theory **57**, 133–141 (2015). https://doi.org/10.1016/j.simpat.2015.06.009
18. Wan, J., Li, L.: Simulation for constrainted optimization of inventory system by using Arena and OptQuest. In: Proceedings—International Conference on Computer Science and Software Engineering CSSE 2008, vol. 2, pp. 202–205 (2008). https://doi.org/10.1109/CSSE.2008.1217
19. Alizadeh, M., Eskandari, H., Sajadifar, S.M.: A simulation model to analyze an invenntory system for deteriorating items with service level constraint. Int. J. Ind. Eng. Theory Appl. Pract. **24**, 587–603 (2017)

Multi-manned Assembly Line Balancing Problem in a Diesel Engine Manufacturing Company: A Real-World Case Study

Leonardo dos Santos Batista and Leandro Magatão

Abstract This article presents a real-world case study for a multi-manned assembly problem issued in an engine manufacturing company: the balancing optimization for a mixed-model production line. In order to describe the real line-balancing problem, the article focuses firstly on the presentation of business needs, and layout and model restrictions assigned. A specific feature called dual station synergy is identified during an early phase of development, which requested mathematical model tailoring, herein described. The use of a mixed-integer linear programming model to solve the problem is presented, as well as its satisfactory results, with reduction of 9% in the cycle time with respect to the current practical solution. As a highlight, this study relied on the physical validation of the proposed mathematical solution in a productive environment of diesel engines.

Keywords Multi-manned assembly line balancing problem (MALBP) · Real-world automotive industry application · Mixed integer linear programing (MILP)

1 Introduction

The automotive industry is known for being a segment that constantly seeks for productivity increase, which often demands large-scale production systems based on assembly lines. In this context, it is commonly necessary to search for reductions in operational cycle time. Many works can be found in the literature for minimizing the cycle time in a line-balancing problem; they are known as Simple Assembly Line Balancing Problem of type 2 (SALBP-2), and contributions to this topic were surveyed by Battaïa and Dolgui [1].

In real cases with a large size product, it is possible for more than one operator to work on the product at the same station, removing the SALBP restriction of only

L. S. Batista · L. Magatão (✉)
Graduate Program in Electrical and Computer Engineering (CPGEI), Federal University of Technology—Paraná (UTFPR), Curitiba, Paraná, Brazil
e-mail: magatao@utfpr.edu.br

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_49

allowing one worker per station and featuring this case as a Multi-manned Assembly Line Balancing Problem (MALBP). This article documents a real case of MALBP in a diesel engine manufacturing company that required a customized solution approach and was optimally solved through an exact technique.

In Sect. 2 the problem is detailed, presenting the company background and the driven business needs. In Sect. 3 the cyclical methodology used to proceed from data collection until optimal solution release is presented. Moreover, the development and tailoring of a Mixed-integer Linear Programming solution is commented. Sect. 4 presents the results of this study in terms of productivity, according to latest released solution. Sect. 5 contains the authors' conclusions along with suggestions for future developments.

2 Problem

2.1 Company Background

The presented problem was identified in a company in the automotive sector, more specifically in a production line for diesel combustion engines for application in heavy vehicles, called herein simply as Line A. Its stations are positioned sequentially in serial straight layout and its cycle time is fixed; it is a paced assembly line [2].

Line A will be upgraded to meet two business needs. The first refers to the increase in production capacity, reflected in a target of reducing cycle time by approximately 15%. The second business need to be accomplished is to adapt current Line A layout to enable the production of a new set of products. This change will modify current settings based as a single model line to a mixed-model line, both characterized by Scholl and Becker [2] and illustrated in Fig. 1.

Currently, the production management implemented for Line A adopts a line balancing optimization process based on the operational experience of engineers

Fig. 1 Assembly lines for single and multiple products [2]

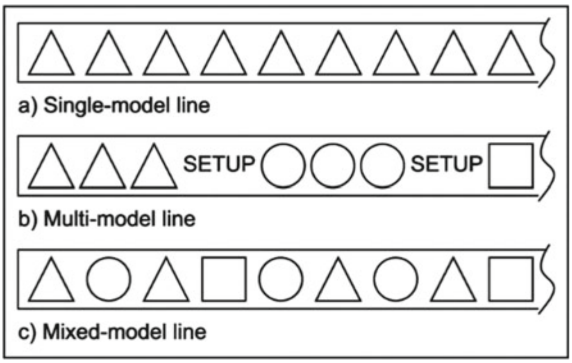


Table 1 Current and future conditions for Line A

Current conditions			Future conditions		
Product model	Production volume (%)	Takt time (unit of time)	Product model	Production volume (%)	Takt time (unit of time)
A	89	1339	A	16	1134
Other models	11		B	53	
			C	5	
			D	21	
			Other models	5	

and technicians who work there. Over time, successive adaptations were made empirically, seeking to optimize the balance for the high runner model.

Considering the future conditions demanded, it is clearly identified the need for an optimized balancing proposal based on future Line A process. However, line balancing changes of this magnitude are hardly executed in an empirical way.

The demanded adaptations of Line A are summarized in Table 1. **Note**: Cycle time and related values, when presented in absolute terms, were multiplied by a common factor, known by the authors, to ensure the company's data confidentiality policy; the values are given in units of time (u.t.). Table 1 indicates that the production *mix* (diversity of different products or production volume of each model) will be affected in future conditions; in current condition, product A accounts for 89% of all the produced combustion engines. In addition, a considerable cycle time reduction is expected (from 1339 to 1134 u.t.), but the complexity of produced models B, C, and D are similar to A product; thus, improvements in, for instance, line balancing and layout are indeed necessary.

2.2 Dual Stations

Line A is currently composed by 11 workstations and 14 operators that operate individually in single stations or in pairs in dual stations, with the freedom to move in a range of positions around the engine. This feature is a characteristic of MALBP, which can be further observed in Michels et al. [3].

Part of the already planned improvement of Line A is the layout adequacy, with the availability of one more workstation and the allocation of one more worker. Setting this way, the line future condition will involve 12 workstations and 15 operators, as showed in Fig. 2. This condition is considered as a baseline for this study. **Note**: Due to physical constraints in the production area, the new station is planned to be positioned in the end of current workstations sequence, as indicated in Fig. 2.

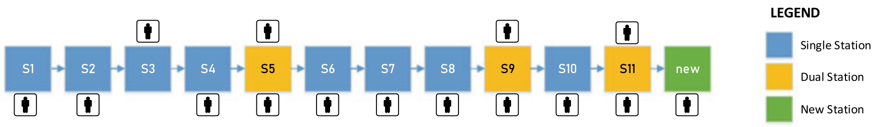


Fig. 2 Line A workstations layout with workers allocation

2.3 Product Positioning

During the initial validation phase of this project, a specific operational condition was found, which needed to be considered in the development of the optimal line balance proposal: the product positioning in dual stations. Due to the large size of the engine with respect to the operators, the product must be adequately positioned for a better access of the operator to the area where each assembly task is realized. This fact contributes, for instance, to reduce fatigue.

This positioning is a semi-automated task (Sikora et al. [4]), done in the product transporter carrier. For stations where only one operator works, this movement happens only once in a simple way to fit ergonomic conditions. However, it becomes more complex when observed in dual stations, due to the need of synergy between tasks that can be done simultaneously in each product side, by two different operators. Setting an optimal product height for a specific assembly in one side of the product may restrict the assembly possibilities in the opposite side, as illustrated in Fig. 3.

The alternative adopted to consider this condition and related customization in the mathematical model is presented in Sect. 3.2. In essence, the mathematical model determines: the tasks to be executed by each operator, in each product side, and for each product positioning. Thus, a series of product positioning decisions in dual stations are modeled. For the best knowledge of authors, no paper has already detailed addressed this feature.

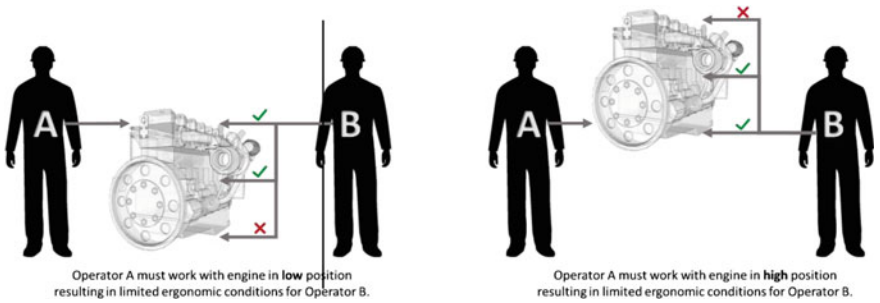


Fig. 3 Product height positioning condition in dual stations

3 Proposed Methodology

3.1 Way of Work Proposed

The data used in real cases of MALBP needs to be constantly reviewed due to some factors that may impact boundary and baseline conditions, such as worker dexterity increase and changes on the production environment. Considering this condition, a cyclical method divided into 5 steps was adopted for conducting this study, which is hereafter detailed:

1. **Data collection**: Identify assembly tasks and sub-tasks, which will be considered in the mathematical model and its respective process time for each product model. The product assembly content is delimited in this step.
2. **Precedence restriction identification**: Set logical and ergonomic constraints that must be considered between tasks to establish a feasible assembly distribution. For instance, the turbo must be assembled before fuel pipes to guarantee correct positioning of both components.
3. **Resource restriction identification**: Set constraints connected to available resources in each station, such as equipment and layout conditions. In this step it is defined the viable reallocations for each assembly task.
4. **Line balance optimization proposal**: Distribute the related tasks in available stations, guided by a specific objective function and according to established restrictions. This step is made with the aid of a mathematical model.
5. **Solution verification**: Virtual or physical verification of the line balance optimization proposal. In this step, the solution feasibility is verified along with the identification of new constraints or necessary adjustments on the mathematical model.

When these 5 steps are concluded, a new balancing version is released. The results that are further presented in Sect. 4 are part of the 6th and the latest balancing solution released, which is provided according to the described methodology.

3.2 Mathematical Model

The proposed mathematical model relies on a Mixed Integer Linear Programming (MILP) approach. The formulation proposed by Sikora et al. [4] was the basis for the model development, but customizations were necessary to include different terms in the objective function, and to correctly represent dual-stations positioning conditions. The objective function (o.f.) main terms are illustrated in Eq. (1):

$$\begin{aligned}\textbf{minimize} z &= k * 10 * Cycle\_Time\_Ref \\ &+ k/10 * MaxDif\_Stations + k/10 * MaxDif\_Workers\end{aligned}$$

$$+ k * sum(\langle t, s\rangle in T_S : \langle t, s\rangle notin TS_{ini}) TS[\langle t, s\rangle]$$
$$+ k * sum(\langle s1, s2\rangle in TwoSideEquil) |STime[s2] - STime[s1]| \tag{1}$$

The o.f. involves the minimization of five terms, weighted by a k factor. The first one is the Cycle Time ($Cycle\_Time\_Ref$), which is a priority to the production system. The second and third involve the maximum temporal difference between stations ($MaxDif\_Stations$) and workers ($MaxDif\_Workers$); these terms are used to try to equilibrate the workload between different stations and workers, even when no more improvements in cycle time are viable. In fact, solutions with equal cycle time (e.g., in case of a bottleneck station), but with different workloads in stations/workers are possible, and these terms contribute to finding better practical solutions in this condition. There is a specific term related to the minimization of changes with respect to the initial line condition, indicated in the TS_{ini} set. Thus, minimizing changes of tasks between stations is also a factor to be considered. The last term is responsible for minimizing the difference between the assembly time in the two sides of dual stations, in different engine positionings.

Basically, the variable $STime[index\_of\_station]$ receives the total time associated to a specific station, which is determined by the sum of all allocated tasks (indicated by the binary variable $TS[\langle t, s\rangle]$) multiplied by the duration of each task. The cycle time is determined by the most loaded workstation ($Cycle\_Time\_Ref \geq STime[s] \forall s \in S$). The variable $MaxDif\_Stations$ is linearly determined as indicated in the set of inequalities (2): the workload of each station is compared to an average workload value (avS), obtained from the equal distribution of workload among all stations. Thus, the maximum difference with respect to avS is obtained. Similar reasoning is used for $MaxDif\_Workers$. For dual stations, a conscious division in adjacent substations is proposed, based on viable product positionings; a balancing in each substation is added to the formulation. For instance, consider the division of S5 in four substations: S5_high_left (*left* side, engine in the *high* positioning), S5_high_right, S5_low_left, and S5_low_right. Specific constraints were added to the formulation to indicate the necessity of equilibrium of workload between, respectively, (S5_high_left and S5_high_right) and (S5_low_left and S5_low_right). Then, the model seeks for solutions that take into account these particular constraints, leading to balancing conditions aligned to the practical needs, even for dual stations. The detailed nomenclature for sets, indexes, parameters, and variables is indicate in Sikora et al. [4], as well as the fundamental constraints involved.

$$\begin{aligned} MaxDif\_Stations &\geq +(STime[s] - avS) \forall s \in S \\ MaxDif\_Stations &\geq -(STime[s] + avS) \forall s \in S \end{aligned} \tag{2}$$

4 Results

The sixth and most recent line balance proposal released is presented in Fig. 4. The experiment was performed in a 64 bit Intel™ i5 CPU—1.6 GHz, 8 GB of RAM, with a time limit of 3600 s. The model was run to optimality within the available time. The generated mathematical model presented 9562 constraints, 4937 variables (4884 binary variables). The results show the total workload in each workstation for each one of the 4 models (A, B, C, and D) that must be produced in de future condition of Line A. The total workload for some stations, and for some specific product models, stays slightly above established target, which in a first look seems to be a non-conformity condition. However, an assumption must be considered for new products assembling: it takes time to workers to achieve the expected assembly dexterity. In this way, a tolerance of 5% in proposed target is adopted for validation purposes. In addition, it can be observed that in each station presenting a model with a time slightly above the planed cycle time (CT), other models have an assembly time inferior to CT. As observed by Lopes et al. [5], in the majority of line balancing cases studied by the authors, it was possible to improve the line productivity by using a sequence aware formulation. Thus, differences of assembly times according to engine models in each station can be consciously used do improve productivity. Such condition can be considered to analyze line-balancing results in combination with product sequence optimization, but are not the focus of this work, which is centered in the already difficult problem of finding an optimized balancing solution, which stills respects a series of practical features.

As a consequence of the proposed line balance optimization, 61 of the total 220 tasks were suggested to be reallocated to different or new workstations. It is important to highlight that the inclusion of station 12 (new station) would naturally imply reallocation of tasks. The proposed changes were checked and the proposed new

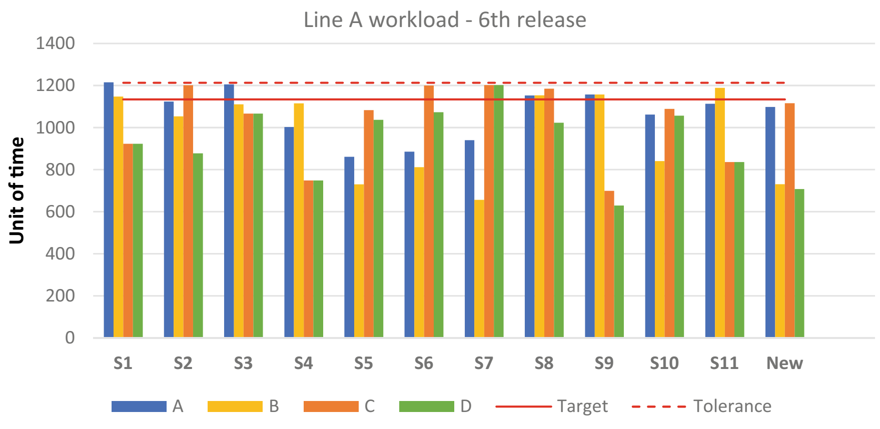


Fig. 4 Proposed line balancing solution – Workload (in time units) for each model in each station for the 6th release

Table 2 Results of balanced workload in dual stations

Workstation	S5		S9		S11	
Workload (unit of time)	928		932		994	
Product position	High	Low	High	Low	High	Low
Worker A	736	187	932	0	977	0
Worker B	693	235	911	0	994	0
Max difference	43	48	21	0	17	0
% of synergy	95	95	98	100	99	100

condition is feasible, considering Line A industrialization planning. Furthermore, the obtained result is in accordance with the defined objectives, modeled by means of Eq. (1), which seeks to primarily reduce the cycle time, and, as a secondary factor, tries to minimize differences with respect to the current balancing practice.

Table 2 presents the workload reached in each dual station for each worker. To check the synergy condition, it is necessary to compare the workload of each worker in both sides and the height product conditions. The less the workload difference between operators A and B the better (the higher) is the synergy; this difference is divided by the station's cycle time to identify the percentage of synergy. Thus, the mixed integer linear programming solution worked efficiently to deal with the MALBP presented in this study, which have special dual-station requirements. The mathematical model tailoring took in account dual station conditions, and resulted as expected: well balanced workload between operators A and B in opposite sides of the product in both, high and low assembly positions.

Another proof of success in dealing with the dual-station condition can noted by the mathematical model decision of avoiding to allocate activities to a given product position, when possible. For instance, in stations 9 and 11 the model indicate the use of high position for the engine, even for the assemblies that are made in the central region (which can be also achieved by the low position). This solution was not achievable in station 5; in it, part of time the engine is assembled in the high position and part in the low one, but maintaining a high synergy of operators, an operational condition that is difficult to be find just based in practical experience.

In addition, it is highlighted that a practical verification was also performed in the presented study: the assembly of 8 units of A-model engines was performed in May 2021, following the 6th release of the line-balancing proposal. The main outcome from physical verification was the validation of the proposed assembly balancing, considering the adherence of workload distribution with values indicated by the model, for all stations. The practical viability of assembling the 200+ involved tasks was physically verified, but the execution of 3 specific tasks would be better managed with precedence constraints not previously informed to the model. This feature tends to be incorporated in a next balancing release.

5 Conclusions

This work addresses a real-world case study for a multi-manned assembly problem in an engine manufacturing company. Special attention was given to modelling of product positioning in dual stations (multi-manned stations), which provided a better adherence with the practical needs. The proposed methodology, described in Sect. 3, relies on solving a Mixed-Integer Linear Programming model, but the physical verification is also a crucial step to validate the solution proposed by the mathematical model, as well as to detect physical restrictions not previously observed, for instance, a precedence link not previously identified. The obtained results suggested a balanced solution aligned to the new target value (as illustrated in Fig. 4); reductions of approximately 9% in cycle time were indicated at the same time that the solution also adequately considered physical positioning conditions in dual stations (as indicated in Table 2). A practical experimentation involving the assembly of a set of A-model engines was performed based on the line-balancing proposal, providing adequate results.

Based on the natural differences of workload in stations, according to the engine model, one possible future development can be the development of a study to optimize the production sequencing, in collaboration with the obtained balancing solution; this synergy can generate increased productivity. Further developments from the theme are exploited in the master's thesis from the first author, available at https://repositorio.utfpr.edu.br/jspui/handle/1/29842 (in Portuguese).

Acknowledgements The authors would like to thank the manufacturing company, in particular the engineer Andrey Lopes for a series of technical analyses, and to CNPq (grant number 311355/2020-0).

References

1. Battaïa, O., Dolgui, A.: A taxonomy of line balancing problems and their solution approaches. Int. J. Prod. Econ. **142**(2), 259–277 (2013). https://doi.org/10.1016/j.ijpe.2012.10.020
2. Scholl, A., Becker, C.: State-of-the-art exact and heuristic solution procedures for simple assembly line balancing. Eur. J. Oper. Res. **168**(3), 666–693 (2006). https://doi.org/10.1016/j.ejor.2004.07.022
3. Michels, A.S., Lopes, T.C., Magatão, L.: An exact method with decomposition techniques and combinatorial Benders' cuts for the type-2 multi-manned assembly line balancing problem. Oper. Res. Perspect. 100163 (2020). https://doi.org/10.1016/j.orp.2020.100163
4. Sikora, C.G.S., Lopes, T.C., Magatão, L.: Traveling worker assembly line (re)balancing problem: model, reduction techniques, and real case studies. Eur. J. Oper. Res. **259**(3), 949–971 (2017). https://doi.org/10.1016/j.ejor.2016.11.027
5. Lopes, T.C., Sikora, C.G.S., Michels, A.S., Magatão, L.: Mixed-model assembly lines balancing with given buffers and product sequence: model, formulation comparisons, and case study. Ann. Oper. Res. **286**, 475–500 (2020). https://doi.org/10.1007/s10479-017-2711-0

Application of Agile Project Management Approaches in the Automotive Industry

Nathalia Juca Monteiro, Robson Luiz Ventura dos Santos, Isabelle Dittert Noleto, Fernando Deschamps, and Sergio E. Gouvea da Costa

Abstract Agile Project Management (APM) is a standard approach in the software industry, and because of its success in allowing projects to better meet its requirements, its application has lately been extended to other industries, without prejudice of its underlying principles and advantages. Among APM practices there are: iterative and incremental deploying the product, presence of informal communication and low bureaucracy and use of visual tools. Additionally, sprints allow the exercise of creativity to solve the presented problem and just-in-time planning to better use people's time. However, adapting practices that were designed for developing a software to other types of products poses many challenges. To shed some light as to how agile practices may be adapted to different contexts, this work conducts a case study of a project in manufacturing that adopted an agile approach. The objective is to analyze how agile practices were applied in this project, from the factory floor to the car dealership. The analyzed project aimed to improve the quality of the after sales services of the company, reducing the percentage of customers problems and complaints related to the service. A semi-structured interview was conducted with the agile coach responsible for the project, which divided the analysis into five areas of study: work and efforts, materiality, agency and creativity, knowledge, and interests and power. Results showed that the highlighted areas are interrelated, since their benefits intertwine while achieving project success. These observed benefits match the application of APM in a software project and were related to practices such as the use of small teams, periodic short meetings, and tools for informal communication. These characteristics made improved project agility and eliminated unnecessary

N. J. Monteiro (✉) · R. L. V. dos Santos · I. D. Noleto · F. Deschamps
Pontifícia Universidade Católica do Paraná, Paraná, Brazil
e-mail: nathalia.monteiro@uepa.br

F. Deschamps
Universidade Federal do Paraná, Paraná, Brazil

S. E. G. da Costa
Universidade Tecnológica Federal do Paraná, Paraná, Brazil

N. J. Monteiro
Universidade do Estado do Pará, Belém, Brazil

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_50

formal documentation and communication. Nevertheless, not all the characteristic tools of APM were used, for example, there was a lack of a project's backlog and the utilization of an agile board, even though visual monitoring was used during the project's execution. Future research is proposed to amplify the study, analyzing other types of projects, and develop a framework with the best practices found to implement the APM methodology outside of software projects.

Keywords Agile project management · Quality · Automotive industry

1 Introduction

The utilization of the Agile Project Management (APM) is an emerging approach that have large acceptance in high-tech enterprises and I.T. software development projects [1–3], but it can be used in other situations [4–6].

The agility can provide some benefits like risk reduction and quality improvement [3]. Stettina and Hörz [7] affirm that despite the APM was originated for software development this methodology received increased attention in the general field of project management. The Agile Manifesto initiated the promotion of the studies, and now the Project Management Institute have a certification called the PMI Agile Certified Practitioner (ACP) [8, 9].

Conforto et al. [1] affirm that the APM approach was developed to improve the performance of the project utilizing agilities methods, tools, and techniques. APM approach differs from the traditional project management approaches because the first one emphasizes continuous design, flexible scope, embracing uncertainty and customer interaction, utilizing a modified project team organization [10].

There is extensive research on the utilization of APM in the software industry and projects, but there is a lack of studies of APM in other industries and projects [11]. The utilization of APM in other areas beyond software development is highly recommended as a response to a fast-changing and challenging business environment [12]. In this scenario, the present study search to answer the following research question: How are the agile tools and techniques applied to project management in an automotive manufacturing company?

The main objective of this article is to provide an overview of how agile tools are utilized for project management in a manufacturing environment. To achieve this goal, a case study was utilized, analyzing the development of a quality project.

This article is divided into six sections. The first presents a contextualization of the work. Section 2 brings a discussion of agile project management. Section 3 discusses the method of the work. Section 4 presents findings from the case study. Section 5 brings discussions related to the findings. Section 6 concludes the article.

2 Agile Project Management

The term "agility" was first utilized in manufacturing and only after it was observed in project management [1]. During the decade of 1980 and early 1990, agility began in project management, focused on software development projects as a response to the challenges experienced in this area [1, 9, 13].

Sweetman and Conboy [14] affirm that during the last 20 years the utilization of agile software project management has been rising. Despite the large utilization of the agile techniques, there is no detailed nor consensual definition among the studies [8, 15], so Conforto et al. [1] proposed a definition of agility utilizing the existent material about the theme. Based on their work the definition is described as follows: "Agility is the project team's ability to quickly change the project plan as a response to customer or stakeholders needs, market or technology demands to achieve better project and product performance in an innovative and dynamic project environment". In summary, when a project faces uncertainty, complexity, volatility, and risk, the agile practices and principles can be utilized [16].

In contrast with the traditional project management approach, the APM focuses on adapting the process and the methodology used to the problem being solved, delivering parts of the project, products, and adjustments as soon as they are requested [12].

The benefits of agile are well known, like an increased ability to manage change priorities, increased productivity, and improved project visibility [14], but it also increases difficulties for the management of the project portfolio [7].

The APM can have a positive impact on a variety of projects, not only in software development [16]. It consists of a highly iterative and incremental process, where there is a close relationship between stakeholders and developers to understand the domain, determine the requirements (what needs to be built), and prioritize functionalities [17]. Due to these characteristics, the APM allows a constant evaluation of the development of the product and immediate feedback from the stakeholders, facilitating modification on the project if necessary [16, 17].

Agile project management is utilized in various situations. Hobbs and Petit [9] examining the APM in large projects in large organizations, concluding that there is still the need for adaptations on the project and organizational levels, but this is an evolutionary process that is still in progress. Scholz et al. [6] applied the APM techniques during the realization of an Industry 4.0 Learning Factory in China, verifying that the results of the tests with customers could be implemented immediately along with the execution of repetitive processes. Mohammed and Karri [5] analyzed agile project management in a construction environment, developing a framework to deal with the dynamic processes.

3 Methodology

The methodology is a case study, which is an empirical study to investigate a phenomenon in a real-life context [18]. This study utilizes the four steps to build a case study proposed by Yin [19]: general planning, data collection, data analysis, and construction of the report.

The general planning included the delimitation of the work and the definition of the company where the study was applied. The industry selected was an automotive company located in Curitiba, Paraná, Brazil. The company is a multinational that has several plants around the globe, and, in Brazil, it manufactures cars. This case study was also chosen, concerning a project created by the manufacturer to improve the quality of the after-sales service, reducing the number of products come backs due to quality issues.

The second phase corresponds to the data collection, which was held through a semi-structured interview, where the authors asked some questions expressed in Table 1 to the project manager to understand the context of the project and how it was developed. The questions were based on the work of Floricel et al. [20], where the theoretical foundations of project management were examined and a toolkit that combined theory and practice in the field of project management was suggested. The toolkit was composed of five areas: work and efforts, materiality, agency and creativity, knowledge, and interests and power. Table 1 resumes the aspect that the authors want to analyze, and the questions asked during the interview.

The data was collected in a conference call with the Agile Coach that has more than 10 years of experience in Product Development and large knowledge in methodologies like Scrum, Lean Projects, and Agile Coaching. This interview lasted one hour and was conducted basically in 2 steps, first a presentation of both parts explaining the objective of the interview, and second, the proposition of the questions and the collection of the answers as free comments by the interviewee. The answers summarized the main ideas of the Agile Coach.

The third phase was the data analysis, where the interview was transcribed and then analyzed to identify the application of APM. In this phase, published literature about APM was utilized to validate or refute the findings. And finally, the article was written.

4 Findings

4.1 Work and Efforts

Due to previous experiences and former knowledge from the interviewee in agile tools and techniques, each area of the study could be verified in the project in question and easily assessed during the interview. Analyzing the first area of study (work and efforts), the interviewee stated that the Agile Team was born in a war room and the

Table 1 Data of the interview

Area	Aspects analyzed	Questions to the interview
Work and efforts	Contextualization of the project, work planning and structure of the team	How is the structure of the project team? The team changed during the project or remained the same? What was the format of the reunions?
Materiality	Instruments of control and reports	How was the project documented? Was there a lot of formal documentation? How were the "deliverables" delivered to the stakeholders?
Agency and creativity	Autonomy and creativity of the team	How much autonomy the project team had? How was the ability of team to adapt to new situations? How were the decisions made?
Knowledge	Share of information between the team and the stakeholders	How was the interaction between the project team? Was there much formalization in the communication between the team? How was the access to the historical data?
Interests and power	Stakeholders and their satisfaction and influence	How much participation and influence the stakeholders had during the project? Were there any conflicts during the execution of the project? How was the project received by the stakeholders? How is the situation of the post-project?

employees were selected from each sector of the company involved in the project. The employees had responsibilities according to the area of work and the power to make decisions during the process chain was encouraged in the project. In this scenario, the project manager's role as a decision-maker was reduced as mentioned by Drury-Grogan [21] and the team members empowered each other to make the decisions by themselves.

The project team must be flexible to team members switch roles to gain new experiences according to Drury-Grogan [21], but in this case, it was not possible since the agile team was not dedicated to the project, meaning their activities were merely a part of their duties. Despite this, the team was cross-functional, complying with one of the characteristics of team members in APM as mentioned by Loiro et al. [4].

As for the meetings, the online environment was preferable (in part due to COVID-19 regulations) and the meetings had a duration of 15 min, happening every day. In addition, an extra one-hour meeting happened weekly to review project directions and the attendance in those meetings were mandatory. However, the meeting's review recommendations in APM are in every two weeks during the execution of the project

[22], which did not extensively comply with the scenario. Nevertheless, the participation of team members at the meetings must be at 72% [23], and the daily meetings are also a vital element of agile practices, therefore, overall, the project has surpassed the ideal conditions.

4.2 Materiality

At the materiality aspect, the agile coach emphasized the need to understand the main necessities of the project in each sprint, looking to solve small goals quickly to add more value during iterations, which can last two to four weeks [5]. The principle of "Simplicity" presented in the Agile Manifesto encourages the condition of continuation if the processes are delivering value to the customer [24].

Since the project involved quality, the project leader confirmed that the deliveries were not necessarily a document, but were also presented as actions, changes, optimizations, or improvements in the process in question. In the analyzed case, the main changes focused on processes developed inside the car dealerships of the brand, and all the evolution of the project was followed in a backlog board.

The project manager described the board as an instrument of control, with items containing tasks to close. The utilization of a board can provides a graphical analysis that can help in the establishment of metrics for the team [25].

The documentation was updated whenever needed in accordance with process modifications or due to incidents. The software Teams was utilized to follow the progress of the activities, updating the progress daily. At the end of the week the team reunited with stakeholders (managers of the company) to present the progress of the agile team.

4.3 Agency and Creativity

About agency and creativity, the team needed autonomy to solve problems and the agile coach had the role of mentoring them. He avoided to point out solutions or preferences in decision making, thus his role was to guide the team to independent improvement. This demonstration of team leader matches the definition proposed by Loiro et al. [4] in APMs, where apart from guidance and support, the team leader also deals with external problems to allow the team to focus on the project. In this scenario, the project leader performs five different activities: facilitating, mentoring, negotiating, protecting, and coordinating [26].

Another important aspect of this studied area is conflict. As the existence of conflicts is inherent to projects, the agile coach emphasized that during the execution of the project, conflict was clearly minimized by the autonomous decision-making done by the team, which reduced friction between members. Drury et al.

[22] also emphasize that teams can make more effective decisions due to the share of knowledge and information.

4.4 Knowledge

In the knowledge area, even though the members of the agile team were not dedicated to the execution of the project, they had roles related to agile activities necessary for the progress of the project, concentrating all expertise inside of the team. It was also stated that they often had informal interactions, held outside of official meetings, which determined project actions. This absence of formalized planning and control approaches is presented by Lappi et al. [24] that affirm agile approaches commonly emphasize informal collaboration.

The agile coach explained that all data was available, and the members of the project team had free access to them, although the information was collected and distributed by the dealerships. It was known inside of the project that shared knowledge within and between teams is one of the key aspects to effective teamwork [27], so the project was described as successful in this area of study, even though the teams that interacted and exchanges information (project and car dealership) worked in different environments and metrics.

Despite the described free access, the agile coach also highlighted some problems in this aspect. Mainly, there was a complex and outdated process to obtain the data from the car dealership team. The use of low technology procedures and manual archives was pointed by the interviewee and the risk of creating doubled reports was identified by the project. This scenario constitutes a problem to the adequate implementation of agile techniques [28].

4.5 Interests and Power

Related to Interests and Power, the literature declares it is a key dimension of measuring project success [29]. Regarding the project, the main interaction between team and stakeholders was to present the result of all the sprints for validation. Even though the agile coach did not mention major problems between the agile team and the final client, he described some resistance from the car dealership's team due to the change of processes. With the support and guidance of the agile coach and the leaders, this resistance was soon contained, and the project continued without issue.

Since the project was not finalized through the company, it is early to measure success, because it only can be quantified after the end of the project [29]. The agile coach emphasized that the project is at the final month of execution, but until now the execution is occurring as planned and the results achieved were expected.

5 Discussions

The challenge for utilizing APM lies in the integration of the agile project at the organizational level [28, 30]. At the project analyzed, an external agile coach was utilized to guide the project team to accomplish its goals.

It is possible to observe that despite the division of areas of study to analyze the utilization of agile techniques, they are interrelated, which means that the aspects analyzed have a strong connection and influence through each other.

The utilization of low technologies was discussed at the Knowledge pilar, and despite the necessity of constant update in this area, the frequent change of technologies may impact the project team performance since the team needs to learn again at every change [27]. This can increase the number of problems and the amount of complexity of the project [27, 31], which can be mainly discussed int the area of Agency and Creativity.

The empowerment of the project team is another subject that was found in more than one area, Work and Efforts, and Agency and Creativity. Team autonomy must be encouraged because it enhances the sense of empowerment [32]. Lee and Xia [33] attribute the team autonomy to the self-organizing nature of the agile team and the interactive process which allows short cycles. To develop self-organization, it is important to develop standards, even if they are small and simple [14].

The close relationship between the agile team and the customers, identified in the Interest and Power area contributes to a better understanding of the project goal and the deliverables expected by the client [24].

Since the interviewee was the project manager, his role at the project was emphasized, and his experience in agile projects could be one of the reasons for the reduction of conflicts between the agile team [26].

It is also important to point out the role of communication in this project. The frequent daily meetings were very adherent with the agile method, allowing objectivity and informal discussions about the progress of the project. In an agile environment, the increased amount of external and internal communication is common [34], so the project analyzed is coherent with the agile characteristics founded in the literature.

6 Conclusions

The APM has large acceptance in software development, but its benefits can also be experienced in other industries and types of projects. The main objective of this study was to analyze how the agile tools and techniques were applied in a quality project of an automotive company.

Thus, the main objective of the study was achieved through the identification of the agile practices during the execution of the project at the vision of the agile coach responsible and the literature described also supported the findings of this research.

The company demonstrated to be familiarized with the concept of agility, so some practices were described as intrinsic to its processes. Despite some project's characteristics not being aligned with common agile practices, the work could expand the scope of agile literature, presenting an APM application outside from the software context.

For future research, the examination of other projects from the same industry is suggested to expand the understanding of agile application in other sectors. It is also valid to interview the stakeholders and the agile project team of the quality project analyzed to observe if the perception of the agile coach was the same to other stakeholders of the project. The development of a framework based on the case study is also recommended to identify best practices and possible guidelines for the implementation of APM outside the software industry.

References

1. Conforto, E.C., Amaral, D.C., da Silva, S.L., Di Felippo, A., Kamikawachi, D.S.L.: The agility construct on project management theory. Int. J. Proj. Manag. **34**, 660–674 (2016). https://doi.org/10.1016/j.ijproman.2016.01.007
2. Lee, S., Yong, H.: Distributed agile: project management in a global environment 204–217 (2010). https://doi.org/10.1007/s10664-009-9119-7
3. Persson, J.S., Mathiassen, L., Aaen, I.: Agile distributed software development: enacting control through media and context **22**(6), 411–433 (2011). https://doi.org/10.1111/j.1365-2575.2011.00390.x
4. Loiro, C., Castro, H., Ávila, P., Cruz-Cunha, M.M., Putnik, G.D., Ferreira, L.: Agile project management: a communicational workflow proposal. Proc. Comput. Sci. **164**, 485–490 (2019). https://doi.org/10.1016/j.procs.2019.12.210
5. Mohammed, K.N., Karri, S.C.: An analytical approach in usage of agile methodologies in construction industries—a case study. Mater. Today Proc. **33**, 475–479 (2020). https://doi.org/10.1016/j.matpr.2020.05.045
6. Scholz, J.A., Sieckmann, F., Kohl, H.: Implementation with agile project management approaches: case study of an industrie 4.0 learning factory in China. Proc. Manuf. 234–239 (2020)
7. Stettina, C.J., Hörz, J.: Agile portfolio management: an empirical perspective on the practice in use. Int. J. Proj. Manag. **33**, 140–152 (2015). https://doi.org/10.1016/j.ijproman.2014.03.008
8. Abrahamsson, P., Conboy, K., Wang, X.: Lots done, more to do: The current state of agile systems development research. Eur. J. Inf. Syst. **18**, 281–284 (2009). https://doi.org/10.1057/ejis.2009.27
9. Hobbs, B., Petit, Y.: Agile methods on large projects in large organizations. Proj. Manag. J. **48**, 3–19 (2017). https://doi.org/10.1177/875697281704800301
10. Serrador, P., Pinto, J.K.: Does agile work?—a quantitative analysis of agile project success. Int. J. Proj. Manag. **33**, 1040–1051 (2015). https://doi.org/10.1016/j.ijproman.2015.01.006
11. Conforto, E.C., Salum, F., Amaral, D.C., Da Silva, S.L., De Almeida, L.F.M.: Can agile project management be adopted by industries other than software development? Proj. Manag. J. **45**, 21–34 (2014). https://doi.org/10.1002/pmj.21410
12. Ciric, D., Lalic, B., Gracanin, D., Tasic, N., Delic, M., Medic, N.: Agile versus traditional approach in project management: strategies, challenges and reasons to introduce agile. Proc. Manuf. 1407–1414 (2019)
13. Eisenhardt, K.M., Tabrizi, B.N.: Accelerating adaptive processes: product innovation in the global computer industry. Adm. Sci. Q. **40**, 84 (1995). https://doi.org/10.2307/2393701

14. Sweetman, R., Conboy, K.: Portfolios of agile projects: a complex adaptive systems' agent perspective. Proj. Manag. J. **49**, 18–38 (2018). https://doi.org/10.1177/8756972818802712
15. Laanti, M., Similä, J., Abrahamsson, P.: What is emphasized in different agile definitions 247–258 (2013). https://doi.org/10.1007/978-3-642-39179-8_22
16. Bergmann, T., Karwowski, W.: Agile project management and project success: a literature review. Adv. Intell. Syst. Comput. **783**, 405–414 (2019). https://doi.org/10.1007/978-3-319-94709-9_39
17. Hass, B.K.: the blending of traditional and agile project management 1–6 (2007)
18. Håkansson, A.: Portal of research methods and methodologies for research projects and degree projects. In: Proceedings of the International Conference on Frontiers in Education: Computer Science and Computer Engineering FECS'13, vol. 13, pp. 67–73 (2013)
19. Yin, R.K.: Cases design and methods. Am. Psychol. Assoc. 1–265 (2003)
20. Floricel, S., Bonneau, C., Aubry, M., Sergi, V.: Extending project management research: insights from social theories. JPMA **32**, 1091–1107 (2014). https://doi.org/10.1016/j.ijproman.2014.02.008
21. Drury-Grogan, M.L.: Performance on agile teams: relating iteration objectives and critical decisions to project management success factors. Inf. Softw. Technol. **56**, 506–515 (2014). https://doi.org/10.1016/j.infsof.2013.11.003
22. Drury, M., Conboy, K., Power, K.: Obstacles to decision making in agile software development teams. J. Syst. Softw. **85**, 1239–1254 (2012). https://doi.org/10.1016/j.jss.2012.01.058
23. Dasí, À., Pedersen, T., Barakat, L.L., Alves, T.R.: Teams and project performance: an ability, motivation, and opportunity approach. Proj. Manag. J. (2020). https://doi.org/10.1177/8756972820953958
24. Lappi, T., Karvonen, T., Lwakatare, L.E., Aaltonen, K., Kuvaja, P.: Toward an improved understanding of agile project governance a systematic literature review. Proj. Manag. J. **49**, 39–63 (2018). https://doi.org/10.1177/8756972818803482
25. Mas, A., Mesquida, A.L., Pacheco, M.: Supporting the deployment of ISO-based project management processes with agile metrics. Comput. Stand. Interfaces **70**, 103405 (2020). https://doi.org/10.1016/j.csi.2019.103405
26. Shastri, Y., Hoda, R., Amor, R.: The role of the project manager in agile software development projects. J. Syst. Softw. **173**, 110871 (2021). https://doi.org/10.1016/j.jss.2020.110871
27. Coakes, E.W., Coakes, J.M., Rosenberg, D.: Co-operative work practices and knowledge sharing issues: a comparison of viewpoints. Int. J. Inf. Manage. **28**, 12–25 (2008). https://doi.org/10.1016/j.ijinfomgt.2007.10.004
28. Jovanović, M., Mas, A., Mesquida, A.L., Lalić, B.: Transition of organizational roles in agile transformation process: a grounded theory approach. J. Syst. Softw. **133**, 174–194 (2017). https://doi.org/10.1016/j.jss.2017.07.008
29. Joslin, R., Müller, R.: The relationship between project governance and project success. Int. J. Proj. Manag. **34**, 613–626 (2016). https://doi.org/10.1016/j.ijproman.2016.01.008
30. Lindvall, M., Muthig, D., Dagnino, A., Wallin, C., Stupperich, M., Kiefer, D., May, J., Kähkönen, T.: Agile software development in large organizations. Computer (Long. Beach. Calif) **37**, 26–34 (2004). https://doi.org/10.1109/MC.2004.231
31. Jarvenpaa, S.L., Shaw, T.R., Staples, D.S.: Toward contextualized theories of trust: the role of trust in global virtual teams. Inf. Syst. Res. **15**, 250–267 (2004). https://doi.org/10.1287/isre.1040.0028
32. Malik, M., Sarwar, S., Orr, S.: Agile practices and performance: examining the role of psychological empowerment. Int. J. Proj. Manag. **39**, 10–20 (2021). https://doi.org/10.1016/j.ijproman.2020.09.002
33. Lee, Xia: Toward agile: an integrated analysis of quantitative and qualitative field data on software development agility. MIS Q. **34**, 87 (2010). https://doi.org/10.2307/20721416
34. Rasnacis, A., Berzisa, S.: Method for adaptation and implementation of agile project management methodology. Proc. Comput. Sci. **104**, 43–50 (2016). https://doi.org/10.1016/j.procs.2017.01.055

An Operational Form of Bundling

Wee Meng Yeo

Abstract This paper introduces a new form of bundling. Our approach is dynamic and contrasts with the conventional notion that bundling should be static over an infinite selling season: One may operationalize two major product line strategies partial mixed-bundling and mixed-bundling dynamically by considering their interaction with firm's inventory information. Our contributions are as follows. First, we develop a model to describe operational bundling and show how one can reduce the complexity of the optimisation problem to obtain joint optimal inventory and bundling strategies. Second, we compute the stationary probability of the joint inventory levels and develop the long-run average payoff rate. Third, we provide necessary and sufficient condition for optimality of operational bundling, which becomes keystone to access the value of static is bundling. Under some conditions, our work shows that bundling strategy should always be dynamic when inventory costs are taken into account.

Keywords Inventory control · Markov chain · Optimization

1 Introduction

We consider a firm selling two items (focal item and non-focal item) where exactly one unit of each type forms a bundle. The firm has two selling options that are differentiated according to various level of sales restriction. The first option is partial mixed-bundling, a practice where the purchase of the focal item must be tied with purchasing the non-focal item. The second option is mixed-bundling, an approach where this tying requirement is removed and the individual focal item becomes available at the original price. We further assume that the firm is able to price the bundle up to the sum of its individual prices. In the atypical situation where there is no bundle discount, mixed bundling is equivalent to unbundling or pure component (see [5]), a practice where each item is available for purchase at its given price.

W. M. Yeo (✉)
University of Glasgow, Glasgow, UK
e-mail: WeeMeng.Yeo@glasgow.ac.uk

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_51

This work sets out to investigate an *operational form of bundling*: adopt mixed bundling when the focal item inventory level belongs to some pre-defined set, otherwise adopt partial mixed-bundling. From a practical standpoint, one may suggest adopting a threshold-based policy where the firm should adopt mixed bundling when the focal item inventory level is above a certain number, otherwise adopt partial mixed-bundling. **How should bundling be operationalized based on inventory information? Is there value to this approach of bundling?** The theoretical reasons for the dominance (if any) of our proposed form of operational bundling is unclear.

This short exposition summarizes our key contributions. **First**, we describe a stylized model for an operational bundling. **Second**, we provide results that allow us to formulate the firm's payoff function. More specifically, we present stationary probability where the firm deploys mixed-bundling and partial mixed-bundling. **Third**, we provide conditions under which our proposed form of dynamic bundling is optimal. We investigate how our proposed form of bundling may outperform mixed-bundling or partial mixed-bundling, selling options which traditionally are assumed to be static.

2 The Model

A firm sells two items (item 1 and item 2). There are two options available: partial-mixed bundling MB_i and mixed bundling MB, generating menus given by $\{\emptyset, \{i\}, \{12\}\}$ and $\{\emptyset, \{1\}, \{2\}, \{12\}\}$ respectively ($\{12\}$ denotes the bundle). Without loss of generality, we assume that $i = 2$ so that item 1 (item 2) becomes the focal (non-focal) item. The firm charges non-negative prices p_i and p for one unit of item i and the bundle $\{12\}$, respectively. Price bundling means that $p \leq p_1 + p_2$ and the bundle price is strictly larger than the focal item price, i.e., $p > p_1$. Therefore, we rule out the case where a manufacturer gladly sells millions of bundles with devices at negative margins or even free, in anticipation of multi-year revenue stream attached to the sales of the consumables. We also assume that p_i is publicly available for consumers' reference within a highly transparent marketplace and is therefore exogenously given. Lastly, we define the variable $\delta = p_1 + p_2 - p \geq 0$ representing the inferred savings from purchasing a bundle, a constraint which we will assume throughout our paper.

The inventory costs are as follows. First, the holding cost is $h_i > 0$ per unit of item i. Second, the marginal cost for each unit of item i is c_i and $c = c_1 + c_2$. Third, the replenishment cost is $K_i > 0$ per replenishment of item i only, and $K = K_1 + K_2$ per joint replenishment of both items. Following the work of [1–3], we assume that replenishment lead-times are negligible, back-ordering is not allowed and there is no benefit to coordinating replenishments.The decision maker applies a zero-inventory ordering policy, i.e., an order is placed only when the inventory level drops to zero, where S_i is the maximum inventory level for item i. We shall define an operational form of bundling as follows.

Definition 1 Let I_i (t) be the inventory level of item i at time t. For any $E \subseteq \Omega_1 = \{1, \ldots, S_1\}$, the firm's bundling regime at time t is:

$$\text{adopt} \begin{cases} \text{partial MB, when } I_1(t) \notin E \\ \text{MB, when } I_1(t) \in E \end{cases}$$

Our definition implies that the firm determines the selling regimes based on set E: partial mixed-bundling (E = ∅) and mixed bundling ($E = \Omega_1$). It turns out that $I_1(t)$ converges in distribution as $t \to \infty$ to a limiting random variable I_1. This convergence allows us to develop the objective function in terms of a stationary payoff rate. The sequence of events is as follows: The firm observes $\mathbf{p} = (p_1, p_2)$, and chooses (i) $\mathbf{S} = (S_1, S_2)$, the order-up-to inventory control policy for each item, (ii) bundle price p, where $p \leq p_1 + p_2$, and (iii) bundling regime $E \subseteq \Omega_1$. Surplus-maximizing consumers arrive according to a Poisson process with rate $\lambda > 0$. By suitably rescaling time, $\lambda = 1$ is assumed without loss of generality. A consumer can be identified heterogeneously by a valuation pair (V_1, V_2) with joint cumulative distribution function denoted by $F(x, y) = \mathrm{P}((V_1, V_2) \leq (x, y))$. Following the work of [4], V_1 and V_2 are assumed to be independent so that $F(x, y) = F_1(x)\ F_2(y)$, where F_i is the continuous marginal distribution function for V_i.

For the bundle, the consumer's valuation is $V = V_1 + V_2$, corresponding to *price bundling** in Stremersch and Tellis [6]. On arrival, the consumer observes the firm's choice of p and E, makes a purchasing decision based on a self-selection mechanism. Define $A_i(y) := \{V_i > y\}$ and $A(y) := \{V > y\}$. Throughout this paper, we denote $A_i = A_i\ (p_i) = \{V_i > p_i\}$, $A = A(p) = \{V > p\}$ and $B_i = A_i\ (p - p_{-i}) = \{V_i > p - p_{-i}\}$. Table 1 summarizes the probabilities that a randomly chosen consumer purchases one unit of item 1, item 2, and bundle, leading to the exact specifications of demand rates to formulate the firm's expected payoff.

For any given $(\mathbf{S}, p, E)$, the firm's expected payoff rate is given by

$$\prod(\mathrm{S}, p, E) := (p_1 - c_1)\mathbb{E}d_1 + (p_2 - c_2)\mathbb{E}d_2 + (p - c)\mathbb{E}d - \sum_{i=1}^{2} (K_i(N_i + N) - h_i \mathbb{E}I_i) \tag{1}$$

Here, d_i and d represent consumers' demand rate for item i *only* and the bundle. Similarly, let N_i be the average ordering frequency of item i *only* and let N be the average ordering frequency of both items together (these quantities are all in stationarity). Because $\lambda = 1$, $\mathrm{E}d_i$ is simply the stationary probability that an arriving

Table 1 Probability that a randomly chosen customer purchases every unit given p and $E \in \{\emptyset, \Omega_1\}$

Bundling regime, E	Item 1	Item 2	Bundle $\{1, 2\}$
∅	0	$\mathrm{P}(A_2 \backslash B_1)$	$\mathrm{P}(A \cap B_1)$
Ω_1	$\mathrm{P}(A_1 \backslash B_2)$	$\mathrm{P}(A_2 \backslash B_1)$	$\mathrm{P}(A \cap B_1 \cap B_2)$

customer purchases item i only, and similarly Ed is the stationary probability that an arriving customer purchases the bundle.

The firm's optimization problem is given by

$$\max_{S,\,E} \prod (S, p, E) \quad \text{s.t. } S = (S_1, S_2),\ E \subseteq \Omega_1, \tag{2}$$

for given p, p_i such that $p_1 < p \leq p_1 + p_2$. Our constraint captures the reality of firm's reduced pricing power due to greater transparency in the online marketplace: the bundle price p is constrained.

3 Analysis

We consider the case where the firm does not offer a bundle discount. This corresponds to solving the optimization problem where $p = p_1 + p_2$. In this case, we have $B_i = A_i$ and $A_1 \cap A_2 \subseteq A$. From Table 1, the probabilities that a randomly chosen consumer purchases every unit of item 1, item 2, and bundle are respectively given by P($A_1 \backslash A_2$), P($A_2 \backslash A_1$), and P($A_1 \cap A_2$) when the firm adopts mixed-bundling. Without a bundle discount, this is equivalent to adopting pure components or unbundling, which we denote by U. The probabilities that a randomly chosen consumer purchases a unit of item 2 and the bundle are given by P($A_2 \backslash A_1$) and P($A \cap A_1$) under partial mixed-bundling.

Let $E \subseteq \Omega_1$ be given and $1_E(m) := 1_{\{m \in E\}}$ be the indicator function. Due to the unit rescaling of our arrival process, the firm's demand rates when $I_1(t) = m$ are given by $d_1(m) = \text{P}(A_1 \backslash A_2) 1_E(m)$, $d_2(m) = \text{P}(A_2 \backslash A_1)$, and $d(m) = \text{P}(A_1 \cap A_2) 1_E(m) + \text{P}(A \cap A_1) 1_{E^c}(m)$. Next, we will determine *derived demand* $b_i := \text{E}(d_i + d)$ for both items. These are the inventory requirements driven by consumer demand. It can be shown that $b_1 = \alpha + r\text{P}(I_1 \in E)$ and $b_2 = \text{P}(A_2) + q\text{P}(I_1 \notin E)$, where $\alpha = \text{P}(A_1 \cap A)$, $r = \text{P}(A_1 \backslash A)$ and $q = \text{P}(A \backslash A_2)$.

As our operational bundling depends on E, it is useful to evaluate the choice between two bundling regimes by considering the joint distribution of two *dependent* stochastic processes $\{I_1(t)\}$ and $\{I_2(t)\}$. It turns out that these two stochastic processes converge in the limit to two independent random variables.

Result 1 ($\delta = 0$). Let S, p and E $\subseteq \Omega_1$ be given. The stationary probability of the joint inventory level $\pi_{(m,n)}$ is given by,

$$\pi_{(m,n)} = \lim_{t \to \infty} \text{Prob}((I_1(t), I_2(t)) = (m, n)) = \text{Prob}(I_1 = m)\text{Prob}(I_2 = n),$$

for some discrete random variable I_1 and $I_2 \sim U\{1, \ldots, S_2\}$.

Result 1 says the following. Under an order-up-to S_i inventory policy, if the firm adopts static bundling, then the stationary joint distribution is modulated by mutually independent I_1 and I_2. In that case, both are uniformly distributed. For non-trivial choice of E, only I_2 is uniformly distributed on the discrete set $\{1, \ldots, S_2\}$.

Result 2 The firm's long-run average payoff is given by

$\prod(p_1, p_2, p, S_1, S_2, E) = \left(p_1 - c_1 - \frac{K_1}{S_1}\right)b_1 + \left(p_2 - c_2 - \frac{K_2}{S_2}\right)b_2 - h_1 \mathrm{E} I_1 - h_2 \mathrm{E} I_2$.

According to Result 2, the payoff for the firm can be expressed in inventory requirement (b_1), rather than demand. It becomes purposeful to answer the following question: What is the optimal form of bundling regime? For this, we will try and solve the optimization problem in (2).

Result 3 Let S_1 and p be given. The optimal subset $E \subseteq \{1, \ldots, S_1\}$ of threshold type. There exists some τ^* such that

$$E = \left\{\tau^* + 1, \tau^* + 2, \cdots, S_1 - 1, S_1\right\}$$

The explanation for Result 3 is that one should deploy partial mixed-bundling when focal inventory level is strictly less than τ or below, otherwise deploy mixed-bundling. This is an intuitive result. Because partial mixed-bundling is restrictive in nature, one should start restricting sales when the inventory level of the focal item falls below τ^*. This is often observed in practice where firms can become risk averse to the availability of the item in providing inventory costs savings for the benefit of operational bundling.

Result 4 Given p and S_i, the firm should adopt the following strategy:

(Static) Mixed-bundling is optimal if and only if

$$h_1 \frac{S_1 - 1}{2} < \frac{\mathrm{Prob}(v_1 > p_1)}{r}(a_1 r - a_2 q)$$

(Static) Partial mixed-bundling is optimal if and only if

$$h_1 \frac{S_1 - 1}{2} < \frac{\mathrm{Prob}(v_1 > p_1) - r}{r}(a_2 q - a_1 r)$$

(Dynamic) Inventory-driven bundling is optimal if and only if

$$h_1 \frac{S_1 - 1}{2} > \max\left(\frac{\mathrm{Prob}(v_1 > p_1) - r}{r}(a_2 q - a_1 r), \frac{\mathrm{Prob}(v_1 > p_1)}{r}(a_1 r - a_2 q)\right)$$

Result 4 presents a necessary and sufficient condition for the dominance of operational bundling over traditional static bundling. It shows that there is benefit to dynamic bundling. Furthermore, if holding costs is zero, then we have the case of informational goods (digital news subscription) and only static bundling holds.

4 Numerical Result

We run an experiment to truly understand the benefit of operational bundling. We assume all forms of replenishment costs to be zero. Let $S_1 = 3, S_2 = 2$ and $(V_1, V_2) \sim U([0, 1] \times [0, 1])$. With either partial mixed-bundling or mixed-bundling, the inventory holding costs for all items incurred turns out to be the *same*. To improve profitability, the firm can switch to partial mixed-bundling regime when the focal item's inventory level falls to one remaining unit, and otherwise adopt mixed-bundling regime, i.e., $E = \{2, 3\}$.

	$p_1 = 0.60, p_2 = 0.30, h_1 = 0.90, h_2 = 0.02, K_1 = K_2 = 0, c_1 = 0.15, c_2 = 0.01$			
	Revenue ($)	Cost 1 ($)	Cost 2 ($)	Profit ($)
Partial mixed-bundling	0.385	0.200	0.03	0.155
Operational bundling	0.384	0.196	0.03	0.158
Mixed bundling	0.383	0.200	0.03	0.153

When $p_1 = 0.60, p_2 = 0.30, h_1 = 0.90, h_2 = 0.02, K_1 = K_2 = 0, c_1 = 0.15, c_2 = 0.01$, operational bundling dominates static regimes because savings accrued from reduced holding costs (of the focal item) outweigh the inevitable loss of revenue.

References

1. Chen, H., Wu, O.Q., Yao, D.D.: On the benefit of inventory-based dynamic pricing strategies. Prod. Oper. Manage. **19**(3), 249–260 (2010)
2. Ignall, E.: Optimal continuous review policies for two product inventory systems with joint set-up costs. Manage. Sci. **15**(5), 278–283 (1969)
3. Liu, L., Yuan, X.-M.: Coordinated replenishments in inventory systems with correlated demands. Eur. J. Oper. Res. **123**(3), 490–503 (2000)
4. McAfee, R.P., McMillan, J., Whinston, M.D.: Multiproduct monopoly, commodity bundling, and correlation of values. Q. J. Econ. **104**(2), 371–383 (1989)
5. Prasad, A., Venkatesh, R., Mahajan, V.: Optimal bundling of technological products with network externality. Manage. Sci. **56**(12), 2224–2236 (2010)
6. Stremersch, S., Tellis, G.J.: Strategic bundling of products and prices: a new synthesis for marketing. J. Mark. **104**, 55–72 (2002)

Multi-objective Simulation Optimization Using Data Envelopment Analysis for Personnel Planning

Fatemeh Jalalvand, Hasan Hüseyin Turan, Sanath Kahagalage, and Sondoss Elsawah

Abstract In this paper, we model and solve a multi-objective personnel planning problem in a military context to find appropriate values of recruitment and promotion policies that simultaneously minimize personnel cost and asset unavailability (due to the personnel shortage to crew assets). We propose a hybrid solution approach that combines a simulation–optimization method with data envelopment analysis (DEA). The simulation–optimization method integrates a genetic algorithm (GA) with a system dynamics (SD) simulation model. While GA can effectively search very large decision spaces, the SD simulation simulates the personnel system and its connections with the fleet of assets. In an interactive process, GA generates solutions of recruitment and promotion policies and sends them to the SD simulation, where the values of personnel cost and asset unavailability are calculated for each solution. Then, DEA computes the efficiency scores of solutions using recruitment and promotion values as inputs, while personnel cost and asset unavailability values as outputs. These efficiency scores are used as fitness values to guide the search process in GA. We test the proposed model on an illustrative example. The numerical results indicate the applicability of the proposed method in identifying the efficient solutions and reducing personnel cost and asset unavailability.

Keyword Personnel planning · Multi-objective · Simulation–optimization · DEA

1 Introduction

Military personnel planning problems usually have multiple objectives that need to be optimized at the same time [1]. In the military context, personnel planning problems attempt to guarantee the right number of the right personnel are recruited and retained to perform organizational tasks such as operationalizing military assets while satisfying personnel budget limitations [2]. On the one hand, military personnel managers tend to minimize personnel cost via strategies including recruiting less personnel and

F. Jalalvand (✉) · H. H. Turan · S. Kahagalage · S. Elsawah
Capability Systems Centre, UNSW Canberra, Canberra, Australia
e-mail: fateme.jalalvand@gmail.com

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_52

decreasing salaries. On the other hand, they need to supply the personnel demand of military assets such as fleet of vessels to crew and operationalize them by adequate and proper personnel and keep these assets available for organizational operations. To this end, military personnel managers should retain reliable amounts of personnel in the organization via strategies such as increasing recruits or increasing salaries to tackle the problem of personnel separation, which imposes unwanted personnel reduction to military organizations. These strategies lead to higher personnel cost. As a result, it is required to simultaneously consider the trade-off between the objectives of personnel cost and asset unavailability.

In this study, we develop and solve a multi-objective optimization problem of military personnel planning to minimize two objectives of personnel cost and asset unavailability, due to the personnel shortages to crew and operationalize assets. To meet the objectives, the recruitment and promotion policies are optimized. Multi-objective optimization problems are often too complex due to, for example, competing objectives and high computational cost [3]. Simulation–optimization methods can analytically address the complexities associated with multi-objective optimization problems [3–5].

In this study, we develop a simulation–optimization method by combining genetic algorithm (GA) and a system dynamics (SD) simulation model. We integrate the developed simulation–optimization method with data envelopment analysis (DEA) to solve the multi-objective optimization problem. DEA models estimate the relative efficiencies of decision-making units (DMUs). A group of studies have attempted to combine DEA and simulation, optimization and simulation–optimization methods. For example, Yun et al. [6] applied DEA to evaluate the fitness of solutions generated by GA in a multi-objective optimization problem. Moreover, Lin et al. [3] developed a discrete-event simulation method coupled with GA and DEA to solve a multi-objective optimization problem in surgical services. They optimized the number of nurses, the number of beds and the labor cost to minimize the objectives of patient waiting time and system completion time. Miranda et al. [7] used DEA to rank the input scenarios of a discrete-event simulation model integrated with an optimization problem in order to identify the best ranges of decision variables and reduce the computational cost. Further, Whittaker et al. [8] developed and solved a bi-level optimization problem in an agri-environmental context using a hybrid GA, which uses DEA to model producer behavior in response to agri-environmental policy. Tsionas [9] developed a formal criterion of weighting based on maximizing proper criteria of optimal model to apply in DEA. In addition, Tsionas [10] considered multi-objective optimization problems from the perspective of generalized DEA and Bayesian analysis to apply in a real-world portfolio optimization problem. Despite the potential of DEA in benefiting the simulation–optimization methods in several ways, some of which mentioned in the above studies, few works, particularly in the Defence field, have integrated DEA with the simulation–optimization methods. This study contributes to close this gap.

The rest of this paper is structured as follows: Sect. 2 explains the proposed multi-objective optimization model. Section 3 describes the simulation–optimization-DEA approach. The computational experiments and results are provided in Sect. 4. At the end, conclusions and future research directions are presented in Sect. 5.

2 Model Description

In this section, we model and present the multi-objective personnel planning problem. The notations for decision variables, sets and parameters applied in the model are provided in Tables 1, 2 and 3.

The first objective of this study is to minimize total personnel cost as the sum of salaries paid to the personnel during the planning horizon T.

Table 1 Decision variables applied in the problem

Decision variable	Description
$x_{i,j,s}^{s'}$	The highest amount of movement of personnel type i in rank j from status $s \in S_{i,j}$ to status $s' \in S_{i,j} \cup S_{i,j+1}$
$y_{i,j,s}^{s'}$	The lowest time required to remain in status $s \in S_{i,j}$ before going to status $s' \in S_{i,j} \cup S_{i,j+1}$. A specific realization of $x_{i,j,s}^{s'}$ and $y_{i,j,s}^{s'}$ is considered as *the promotion* $\mathcal{P}$ *policy*
$z_{i,t}$	The number of recruits for the personnel type i in rank j at time $t \in [0, T]$. A specific realization of $z_{i,t}$ is considered as *the recruitment* $\mathcal{R}$ *policy*

Table 2 Sets applied in the problem

Sets	Description
I	The set of different personnel type in the military
J_i	The set of different ranks for personnel type $i \in I$
$S_{i,j}$	The set of different statuses for personnel type $i \in I$ in rank $j \in J_i$
Ψ	The set of all feasible *promotion* $\mathcal{P}$ and *recruitment* $\mathcal{R}$ *policy*
H	The set of different fleets

Table 3 Parameters applied in the problem

Parameters	Description
T	The length of the planning horizon
$a_{i,j}$	The number of personnel type i in rank j needed to crew any type of asset in a fleet
$E_{t,h}$	The size of fleet $h \epsilon H$ at time $t \in [0, T]$
$d_{i,j,t}$	The salary of personnel type i in rank j at time t

$$\text{Objective 1}: \underset{(\mathcal{R},\mathcal{P})\in\Psi}{\text{minimize}} : \left[\sum_{t=1}^{T}\sum_{i\in I}\sum_{j\in J_i}\sum_{s\in S_{i,j}} d_{i,j,t} \times P^{s}_{i,j,t}(\mathcal{R},\mathcal{P})\right] \quad (1)$$

where $P^{s}_{i,j,t}(\mathcal{R},\mathcal{P})$ denotes the amount of personnel type i in rank j and status s at time t when the recruitment $\mathcal{R}$ and promotion $\mathcal{P}$ policies are taken. Moreover, $d_{i,j,t}$ is the salary of personnel type i in rank j at time t. Moreover, the second objective of our problem is to minimize asset unavailability as the sum of the capacities of vessels that are not crewed formulized in below:

$$\text{Objective 2}: \text{minimize}: \sum_{t=1}^{T}\max\left\{0, \sum_{h\in H} E_{t,h} - V_t(\mathcal{R},\mathcal{P})\right\} \quad (2)$$

where $E_{t,h}$ is the size and composition of fleets $h \in H$ at a particular time t. Furthermore, the maximum number of crewable vessels at time t under promotion $\mathcal{P}$ and recruitment $\mathcal{R}$ policy is $V_t(\mathcal{R},\mathcal{P})$ calculated as follows:

$$V_t(\mathcal{R},\mathcal{P}) = \underset{\forall i\in I, \forall j\in J_i}{\text{minimum}} \frac{P^{s^*}_{i,j,t}(\mathcal{R},\mathcal{P})}{a_{i,j}} t = 1,\ldots,T \quad (3)$$

The constraints on decision variables of the model are provided in Eq. (4). These constraints guarantee that all decision variables are non-negative integers between the related lower and upper bounds.

$$(\mathcal{R},\mathcal{P}) \in \Psi = \left\{\begin{array}{c} x^{s'}_{i,j,s} \le x^{s'}_{i,j,s} \le x^{s'}_{i,j,s} \\ y^{s'}_{i,j,s} \le y^{s'}_{i,j,s} \le y^{s'}_{i,j,s} \\ z_{i,t} \le z_{i,t} \le z_{i,t} \\ x^{s'}_{i,j,s}, y^{s'}_{i,j,s}, z_{i,t} \in \{0\} \cup \mathbb{Z}^{+} \end{array}\right\} \quad (4)$$

The details of the remaining constraints of the problem are provided in [11].

3 Solution Approach

In this study, we mainly follow [3] to develop the solution approach illustrated in Fig. 1, which is a simulation–optimization method integrated with DEA. Similar to [3], we apply GA as an optimization algorithm, however we use SD simulation instead of a discrete-event simulation proposed by [3], which is more compatible to the requirements of our study. Thus, in our solution approach a combination of GA, SD and DEA work together in an iterative process and finally find the set of Pareto optimal solutions with their efficiency scores.

The chromosome structure for the promotion $\mathcal{P}$ and recruitment $\mathcal{R}$ policy of this study is depicted in Fig. 2. In this structure, the sub-chromosome of promotion $\mathcal{P}$

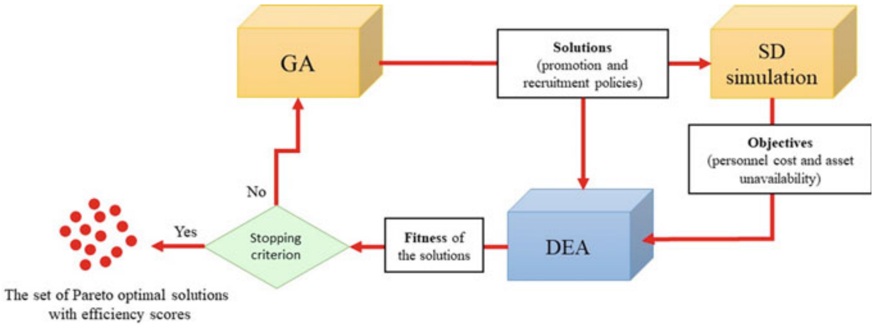


Fig. 1 The developed solution approach

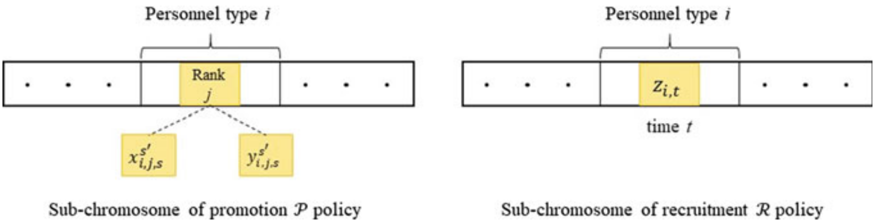


Fig. 2 The chromosome structure of policies

policy corresponds to $x^{s'}_{i,j,s}$ and $y^{s'}_{i,j,s}$, while the sub-chromosome of recruitment $\mathcal{R}$ strategy is related to $z_{i,t}$. The combination of these two sub-chromosomes forms the integrated policy $(\mathcal{P}, \mathcal{R})$.

The tournament selection is applied to choose two solutions as parents. Afterward, two offspring are generated from the two selected parents using a single-point crossover operation on either promotion $\mathcal{P}$ or recruitment $\mathcal{R}$ policy. A mutation operation is also applied by selecting a sub-chromosome and changing the value of a randomly selected gene within it to a random integer between the related lower and upper bounds of variables. The promotion $\mathcal{P}$ and recruitment $\mathcal{R}$ policies generated by these operations create the next generation and are sent to the simulation model.

We set the mutation and crossover probability equal to 0.3 and 0.7, respectively. Moreover, the number of generations is 10, the population size is 120 and the tournament size is 5.

The SD simulation model employs the promotion $\mathcal{P}$ and recruitment $\mathcal{R}$ policies generated by GA to calculate the values of two objectives as personnel cost and asset unavailability by simulating the relationships between the personnel system and the fleet of vessels. The details of the simulation model can be found in [11].

The outcomes of GA and SD are used by DEA to compute the relative efficiencies of solutions (the integrated policies $(\mathcal{P}, \mathcal{R})$) in a population, which are then applied by GA as fitness values to regenerate a new population. The concept of DMU in DEA can be applied here for each solution of the population since we have input values of solutions (promotion $\mathcal{P}$ and recruitment $\mathcal{R}$ policy) and the resulting values of objectives (personnel cost and asset unavailability) as output values for each solution. DEA calculates the relative efficiency of a given DMU in relation to the efficiencies of other DMUs under the assumption that the more efficient DMUs are those that convert lower values of inputs to higher values of outputs [12].

Among different variations of DEA models, we use a constant return-to-scale model presented by Charnes, Cooper and Rhodes (CCR) [12], which is more aligned to the specifications of our study presented in Sect. 4. The linear programming model of CCR is presented in [12], which is applied to calculate the relative efficiencies of DMUs (solutions in our study). Since the values of two objectives (personnel cost and asset unavailability) are undesirable, i.e., the lower values are better, we apply a multiplicative inverse transformation [13, 14] to have a set of desirable outputs. Moreover, since we have the values of decision variables ($x^{s'}_{i,j,s}$, $y^{s'}_{i,j,s}$, $z_{i,t}$) for each quarter of the planning horizon T, we use the average value of decision variables obtained for all quarters of the planning horizon T as the input values of DMUs. According to Ruggiero [15], DEA can properly work with average values of inputs and outputs. The relative efficiency scores obtained by DEA are used to guide GA to regenerate new solutions (promotion $\mathcal{P}$ and recruitment $\mathcal{R}$ policy) for the next generation. After reaching to the stopping criterion (maximum number of generations in GA), the set of Pareto optimal solutions with efficiency scores are identified.

4 Computational Study

To show the applicability of the proposed solution approach, we use a naval personnel planning problem introduced in [11]. In this problem, two personnel types, i.e., officers and sailors, each of which has four ranks ($|I| = 2$; $|J_i| = 4$) are considered. We refer to [11] for details of problem as well as value of input variables.

We choose the planning period for 40 years that covers the years from 2020 to 2060. We run the simulation model with quarterly time steps (i.e., $T = 160$ quarters). During the planning horizon, 6 old vessels are replaced with 12 new vessels according to the fleet renewal strategy applied in the simulation model. GA and DEA were coded in Python 3.7 and the SD simulation was modelled in the AnyLogic 8^{TM}_{ψ} software. We run the proposed approach on an Intel(R) Xeon(R) Platinum 8260 with 12 gigabytes of RAM and 2.40 GHz CPU. The run time was 28,317 CPU seconds.

Figure 3 shows the convergence of the proposed approach to the best solution in terms of the average efficiency score. After generation 4, the average efficiency converges to 1.

During the running of the proposed simulation–optimization-DEA approach, the best 50 solutions are kept and delivered after the termination of the running. We analyze these solutions identified by the approach. The top three efficient solutions in terms of personnel cost and asset unavailability are provided in Tables 4 and 5, respectively.

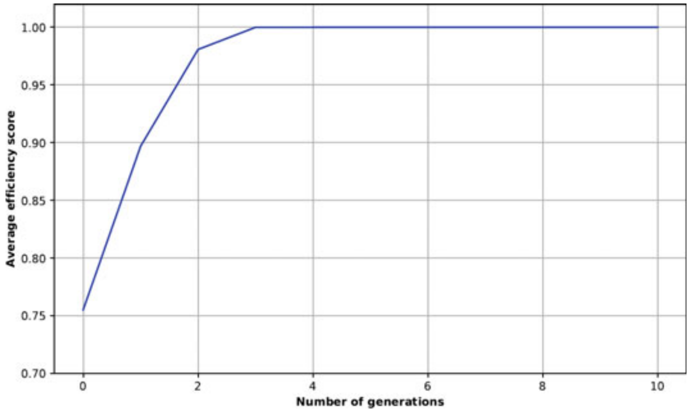


Fig. 3 The convergence of average efficiency of solutions

Tables 4 and 5 show that the lower personnel costs correspond to lower personnel recruitment, especially sailor recruitment. Moreover, while the average amount of personnel movement per quarter is lower, the average time required to remain is higher for the solutions with lower costs (especially for sailors). Furthermore, lower asset unavailability corresponds to higher personnel recruitment, especially sailor recruitment, as shown in Table 5. As expected, more personnel in the system leads to less unavailability of vessels due to the personnel shortage.

In addition, as can be seen in Tables 4 and 5, the least personnel cost (9.35×10^9) happens when asset unavailability is at its third least value (552). Thus, decision makers can take this point as the best efficient solution for the problem.

We analyze the variations of decision variables in efficient solutions using box plots provided in Fig. 4. It is observed that under efficient solutions, we have much higher average time required to remain in a status and lower average amount of personnel movement for sailors compared to officers. This means that under efficient conditions sailors have slower promotions compared to officers, which leads to staying sailors in lower ranks for longer times. This results in lower personnel costs because of keeping sailors in lower ranks for longer time, where they receive lower salaries compared to higher ranks. Since the number of sailors is much higher than that of officers in the organization because of recruiting much higher number of sailors (see Fig. 4c), the strategy of slower promotion of sailors effectively works in reducing personnel cost because lower salaries are paid to personnel that form the majority of the organization's workforce. Moreover, it seems recruiting higher numbers of sailors has reduced asset unavailability due to personnel shortages.

Table 4 Top three efficient solutions sorted by personnel cost

Average recruitment-officers	Average recruitment-sailors	Average amount of personnel movement between statuses-officers	Average amount of personnel movement between statuses-sailors	Average time required to remain in a status-officers	Average time required to remain in a status-sailors	Personnel cost	Asset unavailability
5.79	39.35	10.12	12.93	11.06	11.93	9.35×10^{9}	552
6.25	41.04	9.87	11.06	8.00	10.75	1.13×10^{10}	756
5.79	39.35	9.87	11.50	8.00	13.93	1.14×10^{10}	752

Table 5 Top three efficient solutions sorted by asset unavailability

Average recruitment-officers	Average recruitment-sailors	Average amount of personnel movement between statuses-officers	Average amount of personnel movement between statuses-sailors	Average time required to remain in a status-officers	Average time required to remain in a status-sailors	Personnel cost	Asset unavailability
5.93	43.25	10.12	12.93	11.06	11.93	1.59×10^{10}	548
5.96	43.25	10.12	12.93	11.06	11.93	1.59×10^{10}	548
5.79	39.35	10.12	12.93	11.06	11.93	9.35×10^{9}	552

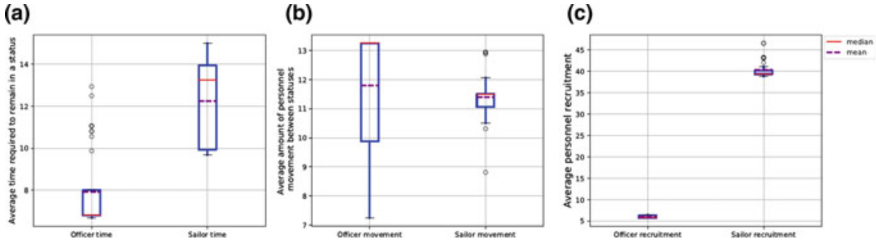


Fig. 4 The variations of decision variables under efficient solutions

5 Conclusion

In this paper, we have addressed a multi-objective personnel planning problem in a military context. We have developed a hybrid-solution approach by combining GA, SD simulation and DEA to solve the problem. We obtained the following conclusions based on our numerical study:

- In efficient solutions, lower personnel costs are achieved by lower sailor recruitment probably because sailors form the majority percentage of workforce in military. Moreover, lower and slower promotions of sailors can be observed for efficient solutions with lower personnel cost.
- In efficient solutions, lower asset unavailability happens when more sailors are recruited.

Our study is limited to using only one fleet renewal strategy, which may not completely reflect the real situations. Future research can develop a dynamic simulation model for the fleet renewal side and combine it with the current simulation model to explore the problem more accurately. Furthermore, we apply constant-return-to-scale DEA model which assumes that an increase in inputs leads to the same proportional increase in outputs. However, further research may focus on applying variable-return-to-scale DEA models to consider non-proportional relationship between the variations of inputs and outputs.

References

1. Jalalvand, F., Turan, H.H., Elsawah, S., Ryan, M.J.: A multi-objective risk-averse workforce planning under uncertainty. In: 2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020, pp. 1626–1633. https://doi.org/10.1109/SSCI47803.2020.9308587
2. Bradshaw, A.E.: United States air force officer manpower planning problem via approximate dynamic programming. Air Force Institute of Technology Wright-Patterson AFB OH Wright-Patterson AFB United States (2016)
3. Lin, R.C., Sir, M.Y., Pasupathy, K.S.: Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: specific application to determining optimal resource

levels in surgical services. Omega **41**(5), 881–892 (2013). https://doi.org/10.1016/j.omega.2012.11.003
4. Rabbani, M., Heidari, R., Yazdanparast, R.: A stochastic multi-period industrial hazardous waste location-routing problem: integrating NSGA-II and Monte Carlo simulation. Eur. J. Oper. Res. **272**(3), 945–961 (2019). https://doi.org/10.1016/j.ejor.2018.07.024
5. Tekin, E., Sabuncuoglu, I.: Simulation optimization: a comprehensive review on theory and applications. IIE Trans. **36**(11), 1067–1081 (2004). https://doi.org/10.1080/07408170490500654
6. Yun, Y.B., Nakayama, H., Arakawa, M.: Multiple criteria decision making with generalized DEA and an aspiration level method. Eur. J. Oper. Res. **158**(3), 697–706 (2004). https://doi.org/10.1016/S0377-2217(03)00375-8
7. de Carvalho Miranda, R., Montevechi, J.A.B., da Silva, A.F., Marins, F.A.S.: Increasing the efficiency in integer simulation optimization: reducing the search space through data envelopment analysis and orthogonal arrays. Eur. J. Oper. Res. **262**(2), 673–681 (2017). https://doi.org/10.1016/j.ejor.2017.04.016
8. Whittaker, G., Färe, R., Grosskopf, S., Barnhart, B., Bostian, M., Mueller-Warrant, G., Griffith, S.: Spatial targeting of agri-environmental policy using bilevel evolutionary optimization. Omega **66**, 15–27 (2017). https://doi.org/10.1016/j.omega.2016.01.007
9. Tsionas, M.G.: Optimal combinations of stochastic frontier and data envelopment analysis models. Eur. J. Oper. Res. (2021). https://doi.org/10.1016/j.ejor.2021.02.003
10. Tsionas, M.G.: Multi-objective optimization using statistical models. Eur. J. Oper. Res. **276**(1), 364–378 (2019). https://doi.org/10.1016/j.ejor.2018.12.042
11. Turan, H.H., Jalalvand, F., Elsawah, S., Ryan, M.J.: A joint problem of strategic workforce planning and fleet renewal: with an application in defense. Eur. J. Oper. Res. (2021). https://doi.org/10.1016/j.ejor.2021.04.010
12. Huang, Z., Susan, X., Zhu, J.: A special issue on "data envelopment analysis: theories and applications" in honor of William W. Cooper. Int. J. Inf. Technol. Decis. Making **4**(3), 311–316 (2005). https://doi.org/10.1016/j.omega.2012.11.003
13. Gomes, E.G., Lins, M.P.E.: Modelling undesirable outputs with zero sum gains data envelopment analysis models. J. Oper. Res. Soc. **59**(5), 616–623 (2008). https://doi.org/10.1057/palgrave.jors.2602384
14. Golany, B., Roll, Y.: An application procedure for dea. Omega **17**(3), 237–50 (1989). https://doi.org/10.1016/0305-0483(89)90029-7
15. Ruggiero, J.: Data envelopment analysis with stochastic data. J. Oper. Res. Soc. **55**(9), 1008–1012 (2004). https://doi.org/10.1057/palgrave.jors.2601779

One-of-a-Kind Productions in Industry 4.0 Environments

Guido Vinci Carlavan and Daniel Alejandro Rossit

Abstract Modern production technologies, based on Industry 4.0, allow to increase the flexibility and resilience of production systems. These enhanced capabilities, along with a greater digitization of the shop-floor, give significant leverage to massive personalization of production. This work addresses an extreme personalization problem, where all products are unique: One-of-a-Kind Production (OKP). OKP problems represent those cases where the production process is focused on the customers' needs and is tailored to their requirements. To address the planning of this type of problems, a CONWIP strategy was used to control production, avoiding overloads in the shop-floor. A particular detail of this work is that the productive configuration adopted is job shop, which is a non-traditional configuration for CONWIP strategy. Furthermore, to better represent the OKP nature of production orders, a great variability was introduced in the number of operations and time required by them. This feature hinders the CONWIP logic, since adding a new job to the shop-floor may not mean adding a workload similar to that of the job that left the shop-floor. To overcome this situation, 2 new dispatching rules that analyze the workload generated by each job before choosing the next job to be dispatched are proposed: SameOp-pure and SameOp-EDD. These rules consider the number of operations of each production order, and prioritize the jobs following a similarity criterion, dispatching as the next job the one most similar to the job that has just left the shop-floor and thus maintaining the workload inside the CONWIP at a fixed level. Computational experiments were carried out based on discrete-event simulation, and the benefits of these new dispatching rules could be verified, obtaining better results than other traditional rules (EDD, FIFO and critical ratio). It was even possible to verify that under more demanding scenarios, the advantages obtained by the SameOp-pure and SameOp-EDD rules were more significant.

G. Vinci Carlavan (✉) · D. Alejandro Rossit
Departamento de Ingeniería, Universidad Nacional del Sur, Bahía Blanca, Argentina
e-mail: gvincicarlavan@gmail.com

D. Alejandro Rossit
INMABB, Universidad Nacional del Sur and CONICET, Bahía Blanca, Argentina

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_53

Keywords One-of-a-kind production · CONWIP · Scheduling · Dispatching rules · Job shop

1 Introduction

The fourth industrial revolution has contributed enormously to the development of flexible production systems with a growing level of autonomy [1]. These capabilities are based on Cyber-Physical Systems (CPS) and Internet of Things (IoT), CPS allows integrating the physical production process with industry information systems in a single system, while IoT allows the different systems and devices to communicate with each other, facilitating joint and collaborative operation of the entire system [2]. In this way, it is possible to integrate the shop-floor with the decision-making processes of the production planning and control systems, which, in turn, have associated intelligent systems that allow the automation of a large part of the decision-making [3]. This increases and improves the capacity and speed of response of the industries, so that production systems are more flexible and resilient to face the changes in customers' demands [4, 5].

These advances in industrial technologies have allowed in recent years a new approach to business models based on mass customization of products [6]. Since, having a highly digitized and interconnected production system allows to significantly improve the handling of the information used and generated by the production system itself, thus facilitating decision-making processes. Precisely, to address highly customized production orders, an agile and efficient information management system is required, because by customizing each production order, standardization is lost and this increases the workload in the Planning and Control Production systems [7]. That is why since the fourth industrial revolution the interest in these customized business models have increased considerably [8].

In this work we address as a production model the extreme customization of the production order, giving rise to One-of-a-Kind production (OKP). In these productions models, the production orders differ substantially from each other, which requires a very flexible production system capable of monitoring unique operation routes for each product [9]. In this sense, Industry 4.0 provides sufficient means for an efficient and successful approach to this type of production, as has already been demonstrated [10]. In their work, Huang et al., use production control loops based on the CONWIP (Constant Work In-progress) logic [11]. CONWIP generates a control loop such that the production system is balanced through "pull" approaches, where it seeks to avoid overloading the production system [12, 13].

The OKP problem is addressed in a job shop productive configuration, where CONWIP is used as a control loop. However, for CONWIP to function properly, a job dispatching rule must be considered in order to select the work to be incorporated into the shop-floor, for that purpose 3 well-known rules FIFO, EDD, Critical Ratio are used, and 2 new rules are also proposed in this work: *SameOP-pure* and *SameOp-EDD*. The two proposed rules seek to level the workload within the shop-floor, not only in number of incorporated jobs (standard function of the CONWIP loop), but also in number of operations. The interesting thing about these new rules is that they allow a more efficient implementation of CONWIP in job shop configurations.

The rest of the work is organized as follows: Sect. 2 gives a description of the problem and the dispatching rules. In Sect. 3 the experiments and computational results are presented, and finally in Sect. 4 the conclusions of the work are introduced.

2 Problem Description

2.1 Model

In order to represent the high variability of OKP demands, ten different product families were proposed, each one having different processing routes, as shown in Fig. 1. Moreover, the probability that an incoming order belongs to a certain family is identic, meaning that the proportion of products belonging to each family tends to be equal. A widely used approach for modeling and addressing CONWIP and OKP problems is Discrete Events Simulation [10, 13], so in this work the same approach will be used.

2.2 Control Loop

The control loop proposed in this work is CONWIP based, meaning that the work in process (WIP) inside the control loop boundaries remains in a constant and predefined level. The governance logic works as follows. When an order is completed and leaves the shop-floor, a RFID device detects the event and sends a signal to the central processing unit of the cyber-physical system. Once the possibility of sending a job to the CONWIP is generated, the need to analyze another question emerges, since it is possible to have multiple jobs waiting in the queue for processing. Thus, it remains to

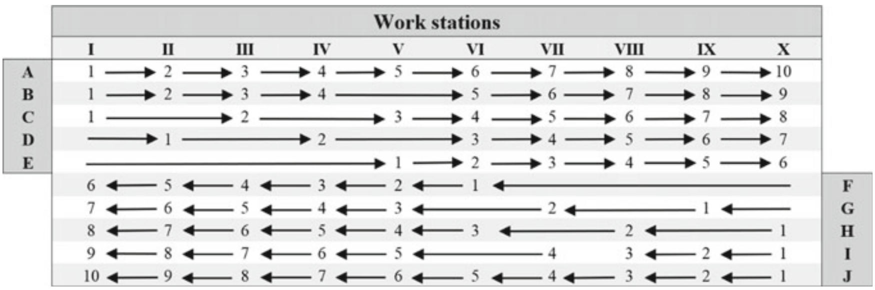


Fig. 1 Number and order of operations for every product family (A, B, C, D, E, F, G, H, I, J)

define which one of these jobs should be selected to go into production. An intrinsic difficulty of OKP systems is that the specifications of a new OKP job order cannot be known in advance, neither the requirements that this order demands from the production system. Therefore, this information is only known once the customer defines the product specifications. This creates extra difficulties for the production planning system, as it must be able to plan orders that are known in real time. To solve this situation, Industry 4.0 provides a very advantageous infrastructure, since, by having all process information digitized and integrated, it is possible to generate intelligent rules that allow to improve the overall performance of the system. It is on this last aspect that dispatching rules are presented in the following sections.

2.3 Dispatching Rules

Since the focus of the work is to generate decision support tools for production planning in OKP environments, it is important to develop methods that allow production planning in an efficient and agile way. In the problems developed in this work, it was considered as an objective function to minimize the number of jobs delivered late. Therefore, it is necessary to define the due-dates of each job *j*, d_j, and when a job can be identified as a late job. Then, it is said that a job is delivered late when its departure date, d_exit_j, is greater than the its due date, i.e. $d\_exit_j > d_j$. So the number of late jobs represents the quantity of jobs that have been delivered after the required due date.

Five different rules were considered to organize the queue that feeds the CONWIP: First In, First Out (FIFO), Earliest Due Date (EDD), Critical Ratio, *SameOp-Pure* and *Same Op and Earliest Due Date* (*SameOp-EDD*).

The FIFO rule is one of the most used rules to organize queuing systems, and it works by attending the arrival order of the jobs, the first job arrived, the first job served. EDD is also widely used for scheduling problems, and seeks that the ordering of the queue represents the priority by proximity to the due date of the respective jobs, giving the highest priority to the job closest to its due date. The *Critical Ratio* is also a well-known sequencing rule, and it is calculated as shown in (1).

$$CR_j = \frac{d_j - t}{\sum_{i=1}^{i=M} \left(p_{ij} + st_{ij}\right)} \tag{1}$$

where CR_j is the Critical Ratio value that job *j* has, d_j is the due date of job *j*, *t* is the time that the simulation run clock marks, st_{ij} is the setup time that job *j* demands on the machine *i* and p_{ij} is the processing time of the job *j* in machine *i*. Therefore, the system sorts in increasing order of CR_j, giving highest priority to the job with the lowest CR_j.

Finally, both of the *SameOp* rules are a proposal from this work. They try to balance the type of work that is inside the CONWIP in order to maintain the level of workload. For this, the *SameOp* family rules seeks to incorporate jobs that are similar to those that have left the job shop. When using CONWIP and either of the *SameOp* as dispatching rule, the control system does not only balance the production in terms of parts (one goes out, one goes in), but also seeks to balance the workload that the part generates on the production system. However, analyzing the workload that each job generates in an OKP job shop production system is not easy, since its nature is high flexibility and many variants. Nevertheless, it is possible to find similar jobs regarding the number of operations that each job requires for their processing, then, *SameOp* searches for jobs l within the queue such that match the number of operations of the job j outgoing from the CONWIP, that is:

$$|O_l| = \left|O_j\right| \tag{2}$$

Then, the rule increases the relative weighting of jobs l that satisfy (2) with respect to the rest of the jobs from the queue. However, it is possible that several jobs in the queue satisfy (2), and because of CONWIP condition (constant WIP) only a single job should be selected. In this case, the *SameOp-Pure* rule selects any job from the queue that satisfies Eq. (2) in a random manner. In contrast, the *SameOp-EDD* rule incorporates Earliest Due Date (EDD) rule as a tie-breaking mechanism in order to establish a secondary priority within the jobs that satisfy (2). As it was explained before, EDD considers the due date of the jobs l, and orders them by priority to those that are closest to being delivered. The algorithm of the *SameOp-EDD* is described with Pseudocode 1.

Pseudocode 1 *SameOp-EDD* dispatching rule.

Input: $Job\_pull\_queue, exit\_job, bonus, time\_left$
Output: job to enter CONWIP

Step 1. $priority_l = time\_left$, for all $l \in Job\_pull\_queue$
Step 2. For $\left\{l : \left|O_{l \in Job\_pull\_queue}\right| = \left|O_{exit\_job}\right|, priority_l = priority_l - bonus\right\}$
Step 3. Order Job_pull_queue in increasing order regarding $priority_l$
Step 4. Pick the first job of the ordering

It is observed that Pseudocode 1 returns as an output from the algorithm, the next job to enter the CONWIP. To solve the problem of which job to select, *SameOp-EDD* starts in Step 1 by assigning the $time\_left$ to all the jobs that are in the queue to enter the CONWIP, where $time\_left$ is the time that each job has left to reach its due date (i.e., $time\_left = d_j - t$). Then in Step 2, the *bonus* is subtracted from those jobs that coincide in number of operations with the outgoing job. The value of the *bonus* is defined in such a way that generates a pre-established difference between those jobs affected by the bonus (i.e., jobs that satisfy Eq. 2), and those that are not. In other words, the bonus parameter determines the relative weight of the *SameOp* and *EDD* components when calculating the priority. Moreover, by adjusting the bonus parameter, the behavior of the *SameOp-EDD* rule can be customized in accordance to the desired performance of the production system. In Step 3, jobs are ordered in increasing priority. It is worth to note that the *bonus* is subtracted from the value already obtained by $time\_left$, therefore, those jobs that had a lower $time\_left$ will be those with the lowest value of $priority_l$, obtaining priority over the rest of the jobs that they have also been affected by the $bonus$. Note that the *SameOp-EDD* rule is still valid even for expired jobs (i.e., $time\_left < 0$), thus, it prevents the case where jobs are indefinitely postponed in the job queue.

3 Experiments and Results

In this section the computational model and studies are presented. To begin with, the experimental design together with the parameterizations and modeling conditions are introduced. Last, the results of the conducted experiments are shown.

3.1 Experimental Design

The processing and setup times for a job n in a workstation m follow a uniform distribution U [1; 1500] min each. This great variability in the determination of the required times, seeks to truly represent the OKP paradigm [9]. The model is formed by ten workstations, each composed by three identical parallel machines. This means that two machines f and g belonging to workstation M1 will perform process operation o_{ij} (where o_{ij} is the operation of job j in workstation $i = 1, \ldots, m$) with the same processing time p_{ij}, therefore, the processing does not depend on which machine within the workstation performs the operation. Also the machines have an uptime rate of 90%, together with a medium time to repair of 5 h.

The order arrival rate follows a negative exponential distribution (similarly to other OKP articles) [10]. The model response was studied under three different scenarios, generated by modifying the arrival distribution mean between 6.5, 7 and 7.25 h, and thus generating three different pressures over the system capacity, from now over referred as exigent, medium and relaxed scenarios.

In order to define the due dates of each job, the total workload required for job j (TWK_j) was considered, affected by a service factor k, as shown in (3).

$$d_j = TWK_j * k + release\_date_j \tag{3}$$

where $release\_date_j$ is the date when job j arrived into the system, and $TWK_j = \sum_{i=1}^{i=m} \left(p_{ij} + st_{ij}\right)$. Therefore, the k factor allows to generate due dates more or less adjusted according to the value that k takes. For instance, in [10], they experimented with k values of 2, 4, 6, 8 and 12, however, the type of productive configuration that they contemplated in their article was a flow shop configuration of 3 stages. In this case a larger flexible job shop type configuration is considered, thus, the value range was expanded to entirely analyze the impact of the k parameter. Therefore, in the present work, $k \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$.

In the performed experiments, each parameter configuration was tested with 50 runs using different generations of random numbers with a confidence level of 95%. On the other hand, by preparatory experimentation, a warm-up period of 2000 days was defined, after which it can be ensured that the system enters a stationary state (meaning that all statistics generated in the first 2000 days are discarded), and data is collected for a duration of 2000 more days, arriving up to a simulation time of 4000 days per experiment.

3.2 Results

In this section, late jobs as percentage of finished orders under the three mentioned scenarios and values of the k parameter are presented, for each of the dispatching rules proposed.

Figure 2a shows the results obtained for the exigent scenario. It can be said that the system is not able to correctly respond to the production demands, since the performance is almost 100% of late jobs for the majority of the rules. However, the only dispatching rule that achieves values significantly different is *SameOp-Pure*, reaching a minimum of 55% for values of $k = 20$. Also, it can be seen that *SameOp-EDD* tends to slightly outperform *EDD, Critical Ratio* and *FIFO* for $k > 16$.

Figure 2b presents the results obtained under the medium scenario. It is possible to appreciate the impact of the arrival rate over the performance of the system, since the late jobs percentage changes significantly compared to the exigent scenario. Also, it can be observed the impact of the k parameter, since the results for $k = 2$ take values of 100% of late jobs for every rule, and for $k = 20$ the vast majority of the rules achieve values around 10%. *SameOp-Pure* has again a better performance, but only for values of $k < 14$, after which *EDD* has the best performance. Also, it can be pointed out that the hybrid rule *SameOp-EDD* generates values that are always between those generated by the rules that conform *SameOp-EDD* (i.e., *SameOp Pure* and *EDD*). Moreover, *Critical Ratio* only outperforms *FIFO* for lower values of k, but for greater values it performs similar to *EDD*. The rule that obtains the worst performance for every value of the k parameter is *FIFO*, being only able to achieve a late job percentage of 40% for $k = 20$, while the best performing rule, *EDD*, results in 0.7%.

Finally, Fig. 2c presents the results obtained under the relaxed scenario. Again, the late jobs percentage is around 100% for $k = 2$, but it decreases in a more abrupt manner when k increases. The rules have similar performance, with the exception of *Critical Ratio* that obtains slightly inferior results for $4 \leq k \leq 8$, and *FIFO*. For $k = 10$, every rule except *FIFO* obtains less than 10% late jobs, and moreover, when $k = 12$, the system produces almost 0% late jobs, or in other words, is able to correctly

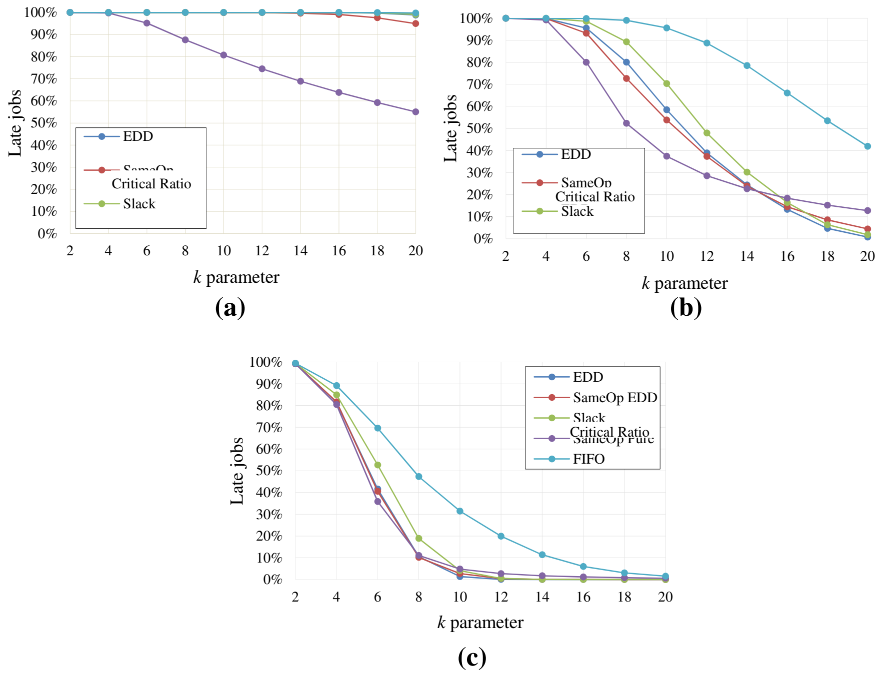


Fig. 2 Late jobs as a percentage of finished orders versus *k* parameter, for each of the dispatching rules under the **a** exigent, **b** medium and **c** relaxed scenarios

respond to all customer demands in time. *FIFO* rule, however, has a much poorer performance achieving values of 1.5% for $k = 20$.

4 Conclusion

This paper addresses a custom production problem of the One-of-a-Kind type in Industry 4.0 production technologies. In order to control the shop-floor of the personalized production, CONWIP was used as a control strategy. However, the performance of CONWIP turned out to be highly dependent on the dispatching rule used to select the next job to be added to the shop-floor. In this sense, it was found that the new dispatching rules proposed in this work allowed a much higher performance than other more traditional rules. The rules proposed here allowed notable improvements, even for the most demanding scenarios in terms of arrival rate of new production orders. On the other hand, there was also a strong dependence on how the delivery date is agreed with the customers, a factor represented by the parameter *k,* being that for higher values of *k* better results were observed in general of the global system.

As future lines of research, it would be interesting to address other peculiarities of the production process, such as the cancellation of production orders that have begun to produce, or to incorporate variability in the actual operating times in relation to those estimated when the order arrives into the system. Also, it would be interesting to further analyze the impact of the bonus parameter as a tool for customizing the hybrid rules.

References

1. Xu, L.D., Xu, E.L., Li, L.: Industry 4.0: state of the art and future trends. Int. J. Prod. Res. **56**(8), 2941–2962 (2018)
2. Lee, J., Bagheri, B., Kao, H.A.: A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manuf. Lett. **3**, 18–23 (2015)
3. Dolgui, A., Ivanov, D., Sethi, S.P., Sokolov, B.: Scheduling in production, supply chain and Industry 4.0 systems by optimal control: fundamentals, state-of-the-art and applications. Int. J. Prod. Res. **57**(2), 411–432 (2019)
4. Ivanov, D., Dolgui, A.: A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. Prod. Planning Control 1–14 (2020)
5. Rossit, D.A., Tohmé, F., Frutos, M.: Industry 4.0: smart scheduling. Int. J. Prod. Res. **57**(12), 3802–3813 (2019)
6. Wang, Y., Ma, H.S., Yang, J.H., Wang, K.S.: Industry 4.0: a way from mass customization to mass personalization production. Adv. Manuf. **5**(4), 311–320 (2017)
7. Rossit, D.A., Tohme, F., Frutos, M.: Production planning and scheduling in cyber-physical production systems: a review. Int. J. Comput. Integr. Manuf. **32**(4–5), 385–395 (2019)
8. Lu, Y., Xu, X., Wang, L.: Smart manufacturing process and system automation—a critical review of the standards and envisioned scenarios. J. Manuf. Syst. **56**, 312–325 (2020)
9. Dean, P.R., Tu, Y.L., Xue, D.: An information system for one-of-a-kind production. Int. J. Prod. Res. **47**(4), 1071–1087 (2009)
10. Huang, G., Chen, J., Khojasteh, Y.: A cyber-physical system deployment based on pull strategies for one-of-a-kind production with limited resources. J. Intell. Manuf. 1–18 (2020)
11. Spearman, M.L., Woodruff, D.L., Hopp, W.J.: CONWIP: a pull alternative to Kanban. Int. J. Prod. Res. **28**(5), 879–894 (1990)
12. Framinan, J.M., González, P.L., Ruiz-Usano, R.: The CONWIP production control system: review and research issues. Prod. Planning Control **14**(3), 255–265 (2003)
13. Grassi, A., Guizzi, G., Santillo, L.C., Vespoli, S.: Assessing the performances of a novel decentralised scheduling approach in Industry 4.0 and cloud manufacturing contexts. Int. J. Prod. Res. 1–20 (2020)

A Novel Graph-Theoretical Approach of Selecting Representative Pareto Optimal Solutions for Multi-objective Optimization Problems

Sanath Kahagalage, Fatemeh Jalalvand, Hasan Hüseyin Turan, and Sondoss El Sawah

Abstract The real-world problems often pose as multi-objective with competing objectives. Unlike single-objective optimization problems, multi-objective problems result in a large set of solutions called Pareto optimal solutions (Pareto set). All the solutions in this set are considered equally good with some trade-offs. Therefore, the decision-makers face the challenge of choosing a solution especially in the absence of subjective or judgmental information. On the other hand, analyzing all the solutions is not practical due to the time complexity. This means that a pruning method is needed to tackle this problem. Several methods have been proposed in the literature. These methods include clustering (e.g., *K*-means) and ranking (e.g., hierarchy process-based) of Pareto optimal solutions to reduce the number of solutions to a promising set with smaller cardinality. In clustering methods, a representative solution is extracted from each cluster to form the reduced set (e.g., the solution at the cluster center or one closest to the ideal solution of the cluster). However, the point closest to the ideal solution may not be a good representation for the entire cluster. Moreover, the reduced set may not contain the extreme solutions and, hence, does not capture the diversity of the entire Pareto set. Therefore, to alleviate the shortcomings of the existing approaches, we propose a novel graph-theoretical approach, which is based on the connectivity (e.g., degree) in the objective space, to obtain the representative solutions from each cluster. We test the applicability of the proposed method on the Pareto optimal solutions obtained from a multi-objective optimization model for a realistic case study. We show both qualitatively and quantitatively that the reduced set obtained from the proposed method better represents the entire Pareto set.

Keywords Multi-objective optimization problems · Pareto optimal solutions · Clustering · Graph theory

S. Kahagalage (✉) · F. Jalalvand · H. H. Turan · S. E. Sawah
Capability Systems Center, University of New South Wales, Canberra, Australia
e-mail: s.kahagalage@adfa.edu.au

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_54

1 Introduction

Single-objective problems are usually not a good representation for many real-world problems as most often they contain many objectives. However, multi-objective problems give a large set of Pareto optimal solutions as opposed to a single solution in single-objective optimization problems. Therefore, the decision-makers face a challenge of choosing a solution especially in the absence of subjective or judgmental information [1]. Thus, finding a reduced set with a small cardinality that captures the diversity of the entire Pareto set is important [2]. This is mainly done via clustering and finding the representative solution for each cluster. Post Pareto analysis with clustering is not new, and researchers have been proposing several methods in the literature (for examples, see [1–4]). There are other existing methods such as hierarchical ranking [4, 5] and data envelopment analysis [6] to obtain the best solution. Nevertheless, these ranking methods depend on the decision maker's preference [4]. On the other hand, what is required for the multi-objective optimization problems are a reduced set of the entire Pareto set which captures the diversity of the solutions.

In clustering methods, a representative solution is chosen for each cluster and the collection of representative solutions for a reduced set. To find the representative solution, the solution at the cluster center or the solution closest to the ideal solution is used [1, 2]. The Silhouette score is used to determine the optimal number of clusters [7]. The solution closest to the ideal solution of the cluster may bias towards the dominant objective function (referring to objective with higher value compared to other values of the objective functions) and, hence, may not be a good representation for the entire cluster if the values of the objective functions are vastly different. The reduced set form from the representative solutions does not include extreme solutions. However, it is important to include extreme solutions to capture the diversity of the Pareto set [8]. Thus, we propose a novel graph-theoretical approach that is capable of finding both the representative solutions for the clusters and extreme solutions. The proposed method facilitates the decision-makers to make their decision with ease. Moreover, the proposed method is independent of subjective decisions and, hence, can be used by a non-experienced decision-maker.

The applications of the post Pareto analysis can be found in the various disciplinaries such as engineering, including energy planning and optimization of renewable energy systems [5, 9, 10], redundancy allocation problem [11], system reliability design problems [3], public emergency service stations [12], and supply-chains [4]. Ayadi et al. [4] stressed that most of the current literature on supply-chain has been focused only on obtaining the Pareto optimal solutions rather than attempting to select a compromise solution. This is also generally true for the multi-objective facility location problems. In a facility location problem, Karatas and Yakıcı [12] proposed a method to find a compromise solution by using branch and bound and iterative goal programming. To the best of the authors' knowledge, there is no post Pareto analysis conducted in the field of military application to obtain a reduced set. The lack of post Pareto analysis in the facility location problems may be due to the fact that most of the facility location problems pose as a single-objective. In particular, around 75% of facility location problems in the military contain a single-objective [13]. To this end, we use Pareto optimal solutions obtained from a military facility location problem to test the applicability of the proposed graph-theoretical approach to achieve a better-reduced set. Thus, we contribute to the military facility location problems by proposing a method to find a better-reduced set and, hence, giving the small Pareto set, which captures the diversity of the entire Pareto set, to the decision-makers.

The rest of the paper is organized as follows. In Sect. 2, we present the proposed graph-theoretical approach to find a reduced set of Pareto set and other relevant background information. Results are presented in Sect. 3 before the concluding remarks in Sect. 4.

2 Methods

In this section, we discuss the proposed method of finding the representative solution for each cluster and extreme solutions. The reduced set then contains both the representative and extreme solutions. To find the representative solutions, first, we group the Pareto set with similar properties together. We do this by using the values of the objective functions in the Pareto set, i.e., in the objective space. K-means clustering with the Silhouette score is used to find the optimal number of clusters within the given upper bound. Then, a graph-theoretical approach which is based on connectivity is utilized to find the representative solution for each cluster and extreme solutions.

The Silhouette score S lies in the range of $[-1, 1]$ and it measures the overall quality of the clustering [14]. It is defined to be the global average of the Silhouette score of individual point $s(i)$. $s(i)$ measures the similarity of a given point i to the other points j in its own cluster (cohesion) compared to the points in the other clusters (separation). If the points in the same clusters are packed together and are well separated from others, then the Silhouette score is high and, hence, so does the quality of clustering. As a general guide, values below 0.2 suggests essentially no clustering pattern was found [2, 15]. The Silhouette score is given by the following equation.

$$S = \frac{1}{n}\sum_{i=1}^{n} s(i) = \frac{1}{n}\sum_{i=1}^{n} \frac{b(i) - a(i)}{\max[b(i), a(i)]}, \quad (1)$$

where $a(i)$ is the average distance in the objective space from the point i representing a Pareto optimal solution to all other points in the same cluster and $b(i)$ is the average of the distances from i to all points in the other clusters.

Once we have the clusters, we find the representative solution for each cluster by using the connectivity of the contact network (graph) created for the Pareto

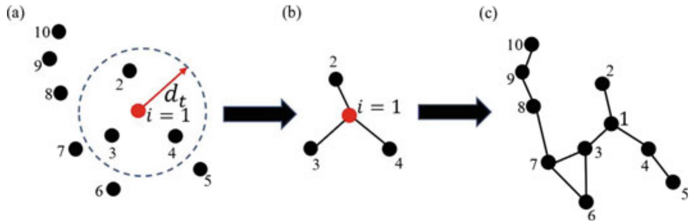


Fig. 1 An illustration of constructing a graph G for the Pareto optimal solutions. **a** We map each Pareto optimal solution to a node. **b** A node $\boldsymbol{i}$ is connected to the other nodes if they lie within a prescribed distance $\boldsymbol{d_t}$ from $\boldsymbol{i}$. **c** This process continues for all the nodes to construct G

optimal solution. Next, we discuss the construction of the graph for the Pareto optimal solutions.

2.1 *Construction of a Contact Network (Graph)* G *for the Pareto Optimal Solutions*

To create a contact network, first, we map each point in the Pareto optimal solutions to a node. For each node i representing a Pareto optimal solution, we connect other nodes if they are within a certain distance $\boldsymbol{d_t}$ in the objective space as illustrated in Fig. 1a and b. The resultant contact network (graph) is denoted by $G = (V, E)$, where V is the set of nodes and E is the set of edges. An edge between node i and j is denoted by e_{ij}. Note that G is an undirected graph for the purpose of this analysis. This means that there is an edge between node i and j if and only if there is an edge between node j and i, i.e., $e_{ij} = e_{ji}$.

2.2 *Proposed Method of Selecting a Representative Solution for Each Cluster*

Our proposed way of finding a representative solution for each cluster utilizes the connectivity of G. We can define some properties of the graph based on the connectivity of the network such as the degree. The degree for a node is defined to be the number of edges incident to that node. For example, the degree of nodes 1 and 7 in Fig. 1c are three as they connect to three other nodes. If a node is connected to many more nodes, its degree is high, and this should be a better representation for the entire cluster. Therefore, the node with a higher degree is used as the representative for that cluster. On the other hand, if the global (referring to the entire graph) degree is very low for node i, that should be a most isolated node and therefore, the corresponding solution sits in corner of the objective space or isolated from other solutions (e.g.,

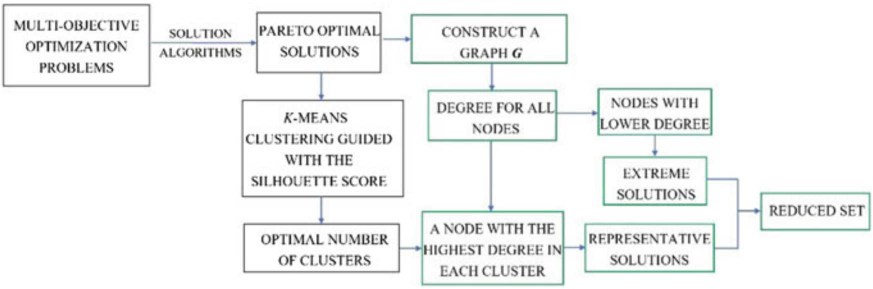


Fig. 2 A flowchart summarizing the proposed graph-theoretical approach to find a better-reduced set that includes both the representative and extreme solutions

nodes 5 and 10 in Fig. 1c. They have the global minimum degree of 1). Thus, we use nodes with lower degrees to identify the extreme solutions. The flow chart summarizing the proposed graph-theoretical approach of obtaining a reduced set is shown in Fig. 2.

3 Case Study

In this section, we present and compare the results from the proposed method with two traditional ways of finding representative solutions, i.e., the solutions closest to the ideal solution and the solution at the cluster center. First, we present the Pareto optimal solutions used to test the performance and the applicability of the proposed novel graph-theoretical method. We solve a multi-objective simulation–optimization problem concerning a realistic case study, which contains three fleet transition options (more details concerning optimization problem, solution approach, and fleet transition options can be found in [16]). We denote these three fleet transition options by $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ and $\boldsymbol{S}_3$. The simulation model is based on a system dynamic (SD) model while the optimization model utilizes the non-dominated sorting genetic algorithm-II (NSGA-II). In Fig. 3, we show the 3D and 2D projections of the entire set of Pareto optimal solutions. Our objectives were to maximize the average fleet availability F_A, minimize the total cost C_T and minimize the total inter-regional transactions I_T. Interregional transactions are defined to be the number of time vessels accessing a dock outside their home-base location. Next, we give the best objective values found. The maximum F_A values are 4.599, 5.197, and 6.021 vessels per week for $\boldsymbol{S}_1 - \boldsymbol{S}_3$, respectively. The minimum C_T values are 1.403×10^{16}, 1.406×10^{16}, and 1.405×10^{16} AUD for $\boldsymbol{S}_1 - \boldsymbol{S}_3$, respectively. The minimum I_T value is zero for each fleet transition option.

In Figs. 4, 5 and 6a and c, we show the representative solution for each cluster with respect to the value of the objective functions for $\boldsymbol{S}_1 - \boldsymbol{S}_3$, respectively. We observed 10, 7, and 5 optimal clusters for $\boldsymbol{S}_1 - \boldsymbol{S}_3$, respectively with the given upper

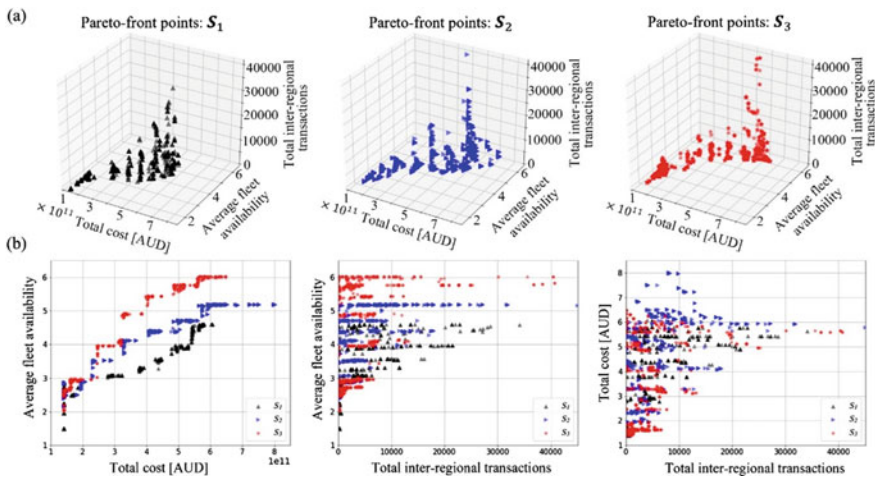


Fig. 3 Visualization of Pareto optimal solutions for three fleet transitions options (S_1, S_2, S_3) in the case study. **a** 3D view. **b** 2D view. We test the proposed method to find a reduced set on these data

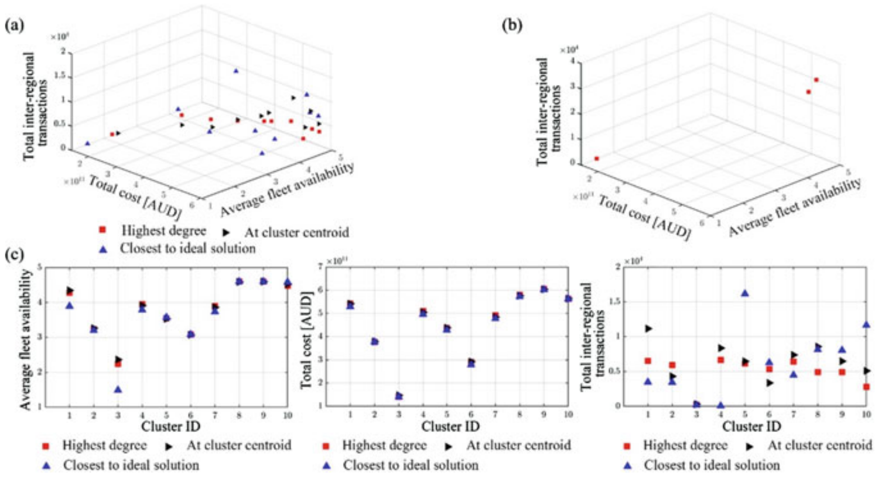


Fig. 4 For fleet transition option 1 S_1. **a**, **b** 3D view. **c** 2D projection of **a** against cluster number. **a**, **c** Representative solutions and **b** extreme solutions. Both the representative and extreme solutions constitute the reduced set

bound of 10. The results depict that the total cost for each representative solution from each method is comparable. Also, we can see that higher F_A and lower I_T for the representative solutions obtained from the solution with the highest degree and the solution at the cluster center compared to the solution closest to the ideal solution. However, representative solutions found by the one closest to the ideal solution result in lower C_T. The solution closest to the ideal solution is biased towards the objective with the highest function value. Therefore, the solution closest to the ideal solution is not a good representation of the entire cluster if the values of the objective functions vastly differ from each other.

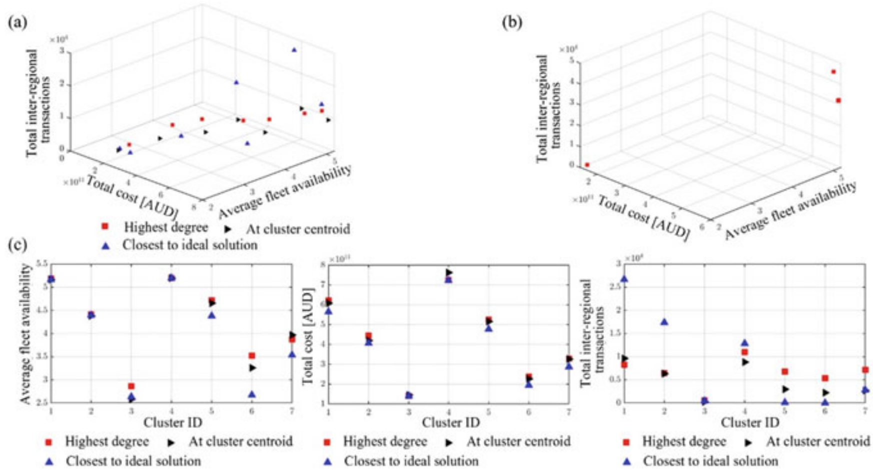


Fig. 5 For fleet transition option 2 S_2. **a**, **b** 3D view. **c** 2D projection of **a** against cluster number. **a**, **c** Representative solutions and **b** extreme solutions. Both the representative and extreme solutions constitute the reduced set

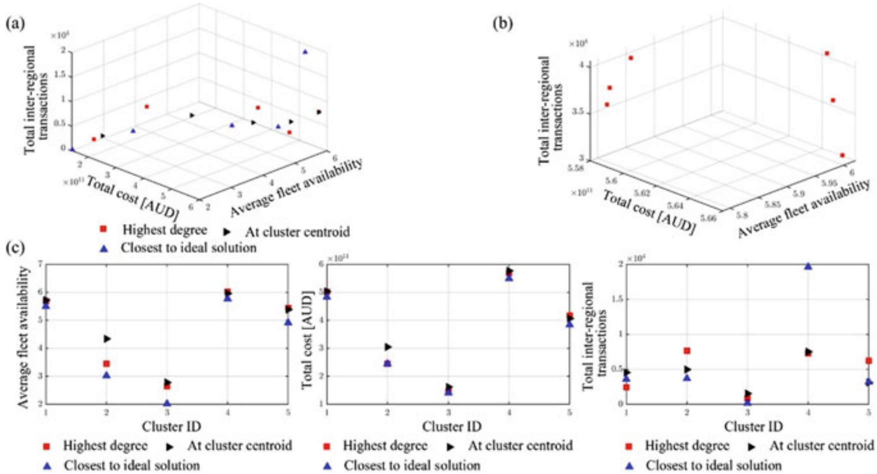


Fig. 6 For fleet transition option 3 S_3. **a**, **b** 3D view. **c** 2D projection of **a** against cluster number. **a**, **c** Representative solutions and **b** extreme solutions. Both the representative and extreme solutions constitute the reduced set

The proposed method can also be utilized to find extreme solutions. As extreme solutions (referring to the maximum or minimum value) at either end of the distribution, we find the solutions with the global minimum degree as discussed in Sect. 2. In particular, to capture other isolated solutions, we used 3% percentile with the left tail of the degree distribution. These extreme solutions are shown in Figs. 4, 5 and 6b for $\boldsymbol{S}_1 - \boldsymbol{S}_3$, respectively. Next, we quantitatively show the importance of including extreme solutions in the reduced set by using the Hypervolume indicator (HI) and the maximum spread (MS). As HI measures the volume of the trade-off space, the higher the HI better the solution is [17]. We used the MS proposed in [18], which compare the Pareto optimal solutions with the ideal and the worst solution. The value of one is desired. In Table 1, we show the values of these indicators for the reduced set obtain from only the representative solutions and both the representative solutions and extreme solutions. Results reveal that the addition of extreme solutions to the representative solutions improves both HI and MS. Therefore, we can conclude that the addition of extreme solutions to the representative solutions gives a better-reduced set, which captures the diversity of the entire Pareto set. Note that, the standalone K-means clustering fails to give extreme solution. However, the proposed graph-theoretical approach can find extreme solutions. Thus, the proposed method is capable of finding the better representative solution and also identifying the extreme solutions. Moreover, the proposed method can be coupled with traditional methods of finding representative solutions (e.g., the solution at the cluster center and the solution closest to the ideal solution) to get a better-reduced set. Hence, the decision-makers get the reduced set with small cardinality, which capture the diversity of the entire Pareto set, to make their decision quickly and accurately with ease.

Table 1 Performance indicators of the reduced set obtained from different methods

Method of selecting the reduced set	Hypervolume indicator (HI)			Maximum spread (MS)		
	$\boldsymbol{S}_1(\times 10^{16})$	$\boldsymbol{S}_2(\times 10^{16})$	$\boldsymbol{S}_3(\times 10^{16})$	$\boldsymbol{S}_1$	$\boldsymbol{S}_2$	$\boldsymbol{S}_3$
Highest degree	2.440	5.510	4.304	0.999	0.884	0.835
Closest to ideal solution	2.085	5.258	3.788	0.993	0.881	0.806
At cluster center	2.520	5.625	4.604	0.982	0.936	0.819
Highest degree + extreme solutions	2.443	5.530	4.306	1	0.884	0.843
Closest to ideal solution + extreme solutions	2.092	5.262	4.649	0.994	0.881	0.852
At cluster center + extreme solutions	2.522	5.642	4.613	0.997	0.941	0.819

4 Conclusions

A novel graph-theoretical approach is proposed to obtain a better-reduced set for multi-objective optimization problems. The reduced set contains both the representative solutions and extreme solutions. We showed that how the proposed graph-theoretical method can be coupled with K-means clustering to obtain a better-reduced set. Our results revealed that this reduced set better represents the entire Pareto set as opposed to reduced sets obtained from the clustering methods alone as proposed in the literature. The proposed method can be utilized with any multi-objective optimization problems. We tested the applicability of the proposed method by using three Pareto sets obtained from a military facility location problem. However, the conclusion drawn here may differ for the other applications and more tests are needed. This is a subject for future research.

References

1. Cheikh, M., Jarboui, B., Loukil, T., Siarry, P.: A method for selecting Pareto optimal solutions in multiobjective optimization. J. Inform. Math. Sci. **2**(1), 51–62 (2010)
2. Zitzler, E., Thiele, L.: An evolutionary algorithm for multiobjective optimization: the strength Pareto approach. TIK-report **43** (1998)
3. Taboada, H.A., Baheranwala, F., Coit, D.W., Wattanapongsakorn, N.: Practical solutions for multi-objective optimization: an application to system reliability design problems. Reliab. Eng. Syst. Saf. **92**(3), 314–322 (2007)
4. Ayadi, O., Felfel, H., Masmoudi, F.: Analytic hierarchy process-based approach for selecting a Pareto-optimal solution of a multi-objective, multi-site supply-chain planning problem. Eng. Optim. **49**(7), 1264–1280 (2017)
5. Antipova, E., Pozo, C., Guillén-Gosálbez, G., Boer, D., Cabeza, L.F., Jiménez, L.: On the use of filters to facilitate the post-optimal analysis of the Pareto solutions in multi-objective optimization. Comput. Chem. Eng. **74**, 48–58 (2015)
6. Friedman, L., Sinuany-Stern, Z.: Combining ranking scales and selecting variables in the DEA context: the case of industrial branches. Comput. Oper. Res. **25**(9), 781–791 (1998)
7. Chaudhari, P., Dharaskar, R., Thakare, V.: Computing the most significant solution from Pareto front obtained in multi-objective evolutionary. Int. J. Adv. Comput. Sci. Appl. **1**, 63–68 (2010)
8. Selvi, S.T., Baskar, S., Rajasekar, S.: Application of evolutionary algorithm for multiobjective transformer design optimization. In: Classical and Recent Aspects of Power System Optimization, pp. 463–504 (2018)
9. Veerappa, V., Letier, E.: Understanding clusters of optimal solutions in multi-objective decision problems. In: 19th International Requirements Engineering Conference, pp. 89–98. IEEE (2011)
10. Abido, M.A.: Multiobjective optimal var dispatch using strength Pareto evolutionary algorithm. In: International Conference on Evolutionary Computation, pp. 730–736. IEEE (2006)
11. Carrillo, V.M., Taboada, H.: A post-Pareto approach for multi-objective decision making using a non-uniform weight generator method. Proc. Comput. Sci. **12**, 116–121 (2012)
12. Karatas, M., Yakıcı, E.: An iterative solution approach to a multi-objective facility location problem. Appl. Soft Comput. **62**, 272–287 (2018)
13. Karatas, M., Yakıcı, E., Razi, N.: Military facility location problems: a brief survey. In: Operations Research for Military Organizations, pp. 1–27. IGI Global (2019)

14. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
15. Tordesillas, A., Kahagalage, S., Campbell, L., Bellett, P., Intrieri, E., Batterham, R.: Spatiotemporal slope stability analytics for failure estimation (SSSAFE): linking radar data to the fundamental dynamics of granular failure. Sci. Rep. **11**(1), 1–18 (2021)
16. Tuan, H.H., Kahagalage, S.D., Jalalvand, F., El Sawah, S.: A multi-objective simulation–optimization for a joint problem of strategic facility location, workforce planning, and capacity allocation: a case study in the Royal Australian Navy. Expert Syst. Appl. **186**, 115751 (2021)
17. Beume, N., Naujoks, B., Emmerich, M.: SMS-EMOA: multiobjective selection based on dominated Hypervolume. Eur. J. Oper. Res. **181**, 1653–1669 (2007)
18. Adra, S.F., Fleming, P.J.: Diversity management in evolutionary many-objective optimization. IEEE Trans. Evol. Comput. **15**, 183–195 (2010)

Graph Model Based Bill of Material Structure for Coupling Product Development and Production Planning

Xiaodu Hu, Adrian Barwasser, Andreas Werner, Frauke Schuseil, Joachim Lentes, Michael Hertwig, and Nikolas Zimmermann

Abstract Currently, a common way of collaboration between product development and production planning is working with a Bill of Material (BOM) in terms of Concurrent Engineering, which is characterized by a structure as a traditional relational data model. Even though such a BOM can serve as a single point of truth of the product structure, it still leads to information discontinuities from upper- to downstream engineering teams, as they use BOM data for different purposes. Hence, a restructuring work of the product development oriented BOM is needed for tasks in production planning, so that the hierarchy is adapted. The parts structure needs to be rearranged to meet the requirements of the manufacturing and assembly processes. In the circumstances of volatile global market and varying customer needs, the BOM structured in relational data model provides insufficient digital consistency to reach an appropriate time-to-market and quick response on product changes and variations. In order to be able to cope with the increasing amounts of heterogeneous data, graph data model approaches are playing an essential role. Due to its non-centric information representation, it can be used as a neutral data model to dissolve boundaries in engineering. A graph model based BOM structure can fulfill both tasks in product development and production planning directly. Thus, it can reach an optimized consistency concerning information flow through various engineering activities as well. This paper proposes the graph model based BOM to minimize efforts and gaps in collaboration between product development and production planning. In addition, it also shows further potentials of usage in engineering.

Keywords Engineering · Product development · Production planning · Bill of Material · Graph data mode · Non-relational data model

X. Hu (✉) · A. Barwasser · A. Werner · F. Schuseil
University of Stuttgart, Institute for Human Factors and Technology Management IAT, Nobelstrasse 12, 70569 Stuttgart, Germany
e-mail: xiaodu.hu@outlook.com

J. Lentes · M. Hertwig · N. Zimmermann
Fraunhofer Institute for Industrial Engineering IAO, Nobelstrasse 12, 70569 Stuttgart, Germany
e-mail: michael.hertwig@iao.fraunhofer.de

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_55

1 Introduction

Manufacturing companies are facing the challenges of data integration during product lifecycle. In respect to engineering processes, Product Data Management (PDM) and Product Lifecycle Management (PLM) systems are applied to manage and share product and engineering data. It is observed that, even though state-of-the-art information management systems are implemented, the collaboration and communication among employees from different business departments still depends strongly on spreadsheets processed manually. For instance, the conventional Bill of Materials (BOM) is initially developed during product development phase and further used by production planning among others. However, due to different purpose of use, manual efforts, e.g. consolidation of various sources of information and restructuring of the BOM are inevitably required to meet the requirements of usage. It leads to information discontinuity throughout engineering processes and consequently more time and effort of communication and coordination. As the global market becomes volatile and customer needs constantly vary, it is essential to ensure time-to-market while keeping costs competitive. This challenge increases the requirements on flexible production, as well as on efficient concurrent engineering as an enabler. As a foundation in engineering, BOM plays an essential role, because it connects product development, production planning and production etc. throughout the entire product lifecycle.

In this context, this paper explores BOM as a starting point to tackle the research question for the collaboration between product development and production planning. Beginning with state-of-the-art review on development history of BOM with related methods and techniques in Sect. 2, the gap is identified. In Sect. 3, a significant workflow change between product development and production planning is introduced as the central research premise of this paper. Next, a graph model based BOM with exemplary data is constructed, along with simple use cases for the purpose of validation. Finally, the potentials of graph model based BOM are discussed in Sect. 4.

2 State of the Art and Qualitative Review

The traditional structure of Bill of Materials (BOM) is cited as "A product structure which defines component relationships consistent with the traditional information requirements of production scheduling and planning …", which is initially used for activities such as product and parts definition and operation scheduling [1]. Traditional structure of BOM have been studied and various structures are developed concerning high-variety product and production, e.g., modular, generative, and generic BOM [2–4]. In addition, researches of BOM structure also addressed collaborative engineering between product development and production, such as the method of compressing Engineering BOM (EBOM) into a three-level Manufacturing BOM (MBOM) to minimize manual trial [5] and a unified Bill-of-Materials-and-Operations (BOMO) to synchronize multiple perspectives on variety [6], etc. Other than direct focus on development and production integration, enterprise-wide methods concerning BOM were widely researched: unified BOM model based on a single source of product data integrating EBOM and MBOM in field of commercial aircraft [7], enterprise BOM in ship design [8], enterprise business information based on Product Structure Management supporting different departments [9] and task-oriented BOM conversion generating multiple views [10], etc. Principally, these former researches are based on relational data model and relational database [11]. Besides, other techniques are also explored for product structure and BOM representation, including Ontology, Extensible Markup Language (XML), etc. [12–14].

In addition to relational approach, the graph theoretic approach was also analyzed and proposed to represent BOM in a tree structure [15–17]. Additionally, tree structured XML model for BOM storage was investigated to improve the efficiency of transforming BOM data from relational model [18]. Furthermore, the approach of graph based BOM with weighted edges has been researched to generate different views [19]. However, tree structures with undirected or weighted edges have limited capacity of representing complex connections between product and production. Nevertheless, a concept of applying complex network theory for Advanced Manufacturing System (AMS) has been studied [20]. It inspires the usage of weighted direct graph to represent product structure using nodes and relations. Subsequently, procedures of converting single and multiple BOMs into networks are proposed with the results in gaining new viewpoints of content and insights into parts [21]. Meanwhile, approaches of graph-based BOM modelling were also developed to analyze complex product structures, in order to enhance value-adding variety concerning standardization and modularization [22].

In summary, various approaches and methods have been researched regarding BOM modelling and representation. Traditional relational data models play an important role given that numerous mature PLM applications are based on that on the market. However, relational models face the challenges when handling complexly networked product and production data. As an emerging technology, graph database receives increasingly attention in industry and institutions [23] and shows great potentials to deal with networked data. Thus, this paper addresses especially the introduction of graph model based BOM to tackle the networked product and production data for better coupling of product development and production planning.

3 Graph Model Based BOM

3.1 Changes in Engineering Workflows

This paper is based on the premise that the introduction of graph database for BOM can increase the collaboration efficiency between product development and production planning and result in less information discontinuities. Figure 1 illustrates the premise: Changes of workflow in engineering by introducing graph database for BOM structuring and storage.

As depicted in (a) of Fig. 1, via the relational model approach, engineers develop BOMs mostly motivated by constructing parts and subparts to fulfill the required functions of a product. With additional engineering information e.g. part number, variants and department information, the EBOM is constructed in a way that each part is assigned to a hierarchy level representing their functional and structural relation to others. However, the EBOM need to be enriched with production information to serve the production planning. Therefore, a mapping or restructuring process from EBOM to MBOM is required. The common approach is manual adaption of BOMs e.g. via spreadsheets, where the efficiency and correctness of adapting depend highly on the experience and expertise of the planner. Besides, PLM software provides the function to transform EBOM to MBOM through an adaptor. Nevertheless, manual efforts for re-examination, correction and maintenance are inevitable in practice. Furthermore, human errors, incompleteness of information and frequent changes in BOM related engineering activities are accumulated during forward propagation, thus not being responded in time.

The proposed approach is aiming at resolving the problems mentioned above by introducing the graph model based BOM, as shown in (b) of Fig. 1. On one hand,

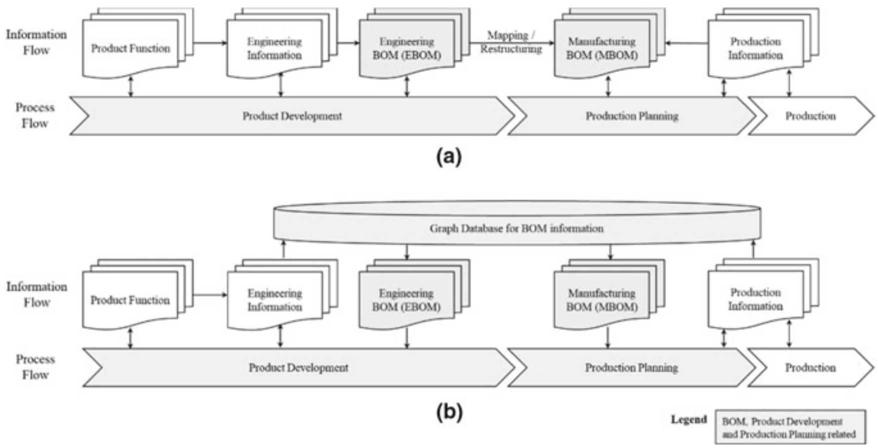


Fig. 1 Workflow changes in collaboration between product development and production planning using **a** relational data model based BOM to **b** graph model based BOM

graph model is commonly considered as a well-performed approach of knowledge representation, which lowers the high demand on planner's experience and expertise at elaborating MBOM; on the other hand, the graph database for BOM information storage is used as the single point of truth for both product development and production planning. It can alleviate the effect of error accumulation and information discontinuity along with long process chain.

3.2 Exemplary EBOM and MBOM

To setup the research, this paper takes an exemplary dataset of souvenir clock from [6] as reference. The dataset includes:

Bill-of-Materials (BOM) data for engineering: It describes the hierarchy level, parent item, component item of souvenir clock, etc.
Bill-of-Operations (BOO) for production: It contains the operation sequence number, operation (name), work center, etc.
Bill-of-Materials-and-Operations (BOMO) data for production planning: Based on BOO, corresponding material (component item) for operation and respective output (parent item) etc. are included.

The aforementioned dataset is then adapted for the purpose of applying graph database. The adapted dataset are summarized in Tables 1 and 2. In practice, components and parts are managed by unique part number in industry. Parts are strictly differed from each other using part number, although some parts share the identical name. For example, the part Label Stick (Table 1) has two part number, because the part has different state during production. Thus, to represent each item and operation as unique node in graph database, each item (part) from the referenced dataset is assigned with a dummy unique identifier under the header *Item UID* in Table 1, while the header *Work Center* from the original BOO as operation reference is redefined as *Operation UID* in Table 2. As result, either part or operation items can be referenced by another line (e.g. *Parent Item* in Table 1 and *Material* in Table 2) are replaced by their unique identifier. In addition, the EBOM (Table 1) is extended with an item Raw Material, which is part of the production but not hierarchized into product.

3.3 Graph Data Model of BOM

This paper utilizes the tool Neo4j as graph database with its Cypher language for query [24]. The query result is visualized through an intuitive user-interface with nodes and relations. The modelling of the graph model based BOM is realized via mapping the exemplary EBOM and MBOM of the souvenir clock in Python scripts with the python-library py2neo.

Table 1 Exemplary EBOM of souvenir clock

Hierarchy level	Item UID	Item name	Parent item	Quantity
1	SC0001	Souvenir clock		1
2	DC0001	Desk clock	SC0001	1
3	BO0001	Body	DC0001	1
4	HA0001	Hands	BO0001	1
4	DI0001	Dial	BO0001	1
4	SP0001	Spacer	BO0001	1
4	MO0001	Movement	BO0001	1
5	GS0001	Gear Set	MO0001	1
5	TR0001	Transmission	MO0001	1
5	CO0001	Core	MO0001	1
5	CA0001	Case	MO0001	1
5	CV0001	Cover	MO0001	1
4	SR0001	Screw	BO0001	4
3	FA0001	Frame	DC0001	1
4	BA0001	Base	FA0001	1
4	FP0001	Front plate	FA0001	1
4	LS0002	Label sticker	FA0001	1
2	PB0001	Paper box	SC0001	1
2	LS0001	Label sticker	SC0001	1
N.A.[a]	RM0001	Raw material	N.A.[a]	3

[a]*N.A.* not available

Graph data model consists of two basic components: nodes and relations. Nodes in graph data model possess one or multiple labels and properties. Labels enable an efficient sorting and filtering of nodes, while properties contain all information. Relations in graph data model connect two nodes and must have a direction. One relation has one or multiple labels and properties.

In the initial step, corresponding graph data model is designed based on the exemplary EBOM and MBOM of souvenir clock. As shown in Table 3, two different node labels are defined. The label *part* and *operation* cover all items in Table 1 and all operations in Table 2. Each node labeled as *part* has the properties *UID*, *name* and *parentItem*, which correspond to the part number, part name and superordinate part. Meanwhile, the properties *UID*, *name*, *type* and *nextOP* in the nodes labeled as *operation* represent operation work station code, operation name, operation type and the next operation.

In the following step, four relations are defined in Table 4. The relation *parent* describes the hierarchical affiliation between parts, but does not include the cross-level representation among all parts. The relation *subsequent* depicts the topology of operations and determines whether an operation is the next step from the last

Table 2 Exemplary MBOM of souvenir clock

Operation name	Operation UID	Material	Product
Packaging and inspection (A5)	WC-A5	DC0001	SC0001
Kitting (K7)	WC-K7	LS0001, PB0001	
Clock assembly (A4)	WC-A4	HA0001	DC0001
Kitting (K6)	WC-K6	SP0001, MO0001, SR0001	
Paper box preparation (A2)	WC-A2	RM0001	PB0001
Kitting (K4)	WC-K4		
Frame assembly (A3)	WC-A3	BA0001	FA0001
Kitting (K5)	WC-K5	FP0001, LS0002	
Movement assembly (A1)	WC-A1	GS0001	MO0001
Kitting (K3)	WC-K3	TR0001, CO0001, CA0001, CV0001	
Base fabrication (M1)	WC-M1	RM0001	BA0001
Front plat fabrication (M2)	WC-M2		FP0001
Kitting (K1)	WC-K1		
Printing (M3)	WC-M3	RM0001	LS0002
Kitting (K2)	WC-K2		DI0001

one. For instance, the parallel execution of operations is not considered. The relation *material_to* shows the relationship between parts or materials and operation and indicates the required parts or materials for an operation, while *produce* refers to the output part or product of an operation. Both of them are interrelation between part and operation nodes. Figure 2 illustrates the modeled data in graph database with the defined nodes and relations above.

Figure 2a presents all stored nodes labeled as *part* and their relations. One exception is the single node without any connection to other parts, because the node of raw material has neither sub- nor superordinate parts. This illustration shows intuitively the hierarchy of all components and parts of souvenir clock and provides the basic view of EBOM in a proper way to satisfy engineers in product development.

Figure 2b depicts the nodes labeled as operation and their relations. It gives a clear view of the sequence of all operations in production, which supports the production planning in process planning, etc.

Figure 2c shows the holistic representation for all stored data with their interrelations. It is the pursuit data model in this paper—an integrated and interrelation storage of both engineering and manufacturing BOM as a graph model based BOM for an optimized collaboration between product development and production planning.

Table 3 Exemplary node label and property in graph model based BOM

Node label	Node property	Description
part		The label *part* is explicitly set to differ from the label operation. It represents all type of items e.g. raw materials, components, sub parts or final product, regardless of their hierarchy level
	UID	The property *UID* is the unique identifier for identification of the part, as well as for generation of relations within and between parts and operations
	name	The property *name* as human-understandable description of one part is used for visualization in human–machine-interfaces
	parentItem	The property *parentItem* contains the UID of any superordinate part. It defines the affiliation among parts with regard to the hierarchy level
operation		The label *operation* is explicitly set to differ from the label part. It represents all type of operations e.g. kitting, assembly and manufacturing, regardless of their sequence
	UID	The property *UID* is the unique identifier for identification of the operation, as well as for generation of relations within and between parts and operations
	name	The property *name* provides human-understandable text description of one operation and is used for visualization in human–machine-interfaces
	type	The property *type* describes the operation type e.g. assembly. It defines the responsibility within production planning department
	nextOP	The property *nextOP* contains the *UID* of next operation. It defines the sequence of production operations

Table 4 Exemplary nodes' relation in graph model based BOM

Relation	Description
parent	The relation *parent* represents the hierarchy of parts
subsequent	The relation *subsequent* describes the process order
material_to	The relation *material_to* shows, which part or materials are to be processed in an operation
produce	The relation *produce* determines, which part is produced by an operation

3.4 Validation with Use Cases

To investigate the feasibility of graph model based BOM approach for an optimized coupling of product development and production planning, as well as the fulfillment of activities for both sides, four use cases are defined (Table 5).

Case A and Case D describe the scenario that engineers or planners need information from BOM only involving their own functional area. In these two cases, no cross-departmental collaboration is required. Case B and C depict the situation

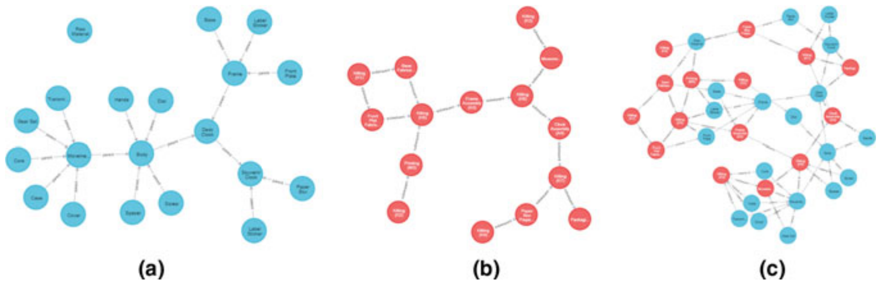


Fig. 2 Visualization of nodes and relations **a** between parts, **b** between operations and **c** among parts and operations

Table 5 Use cases of engineering activities within product development and production planning

	Involving product development	Involving production planning
Engineer (product developer)	**Case A**: Engineer wants to view all parts and their structure to support the design of product variants	**Case B**: Engineer wants to see the process information concerning his responsible part, as well as the belonging subparts, in order to apply design-for-manufacturing/assembly rules in construction
Planner (production planner)	**Case C**: Planner wants to see the materials and parts to be processed and manufactured in his/her responsible production operation, in order to confirm the completeness of jigs and tools around the work station	**Case D**: Planner wants to view all existing production operations and their sequence for this product, in order to determine whether another variant in the same product family can be produced without any changes

that engineering or planner requires information, which are beyond their responsible areas.

For each use case, Cypher language is used upon the graph model based BOM to query and regarding the use cases. Figure 3 shows the query examples and the corresponding result representation.

For Case A and Case D, as mentioned for Fig. 2, the topology with directed relation contains the characteristics of traditional BOM inherently. Thus, through simple Cypher query specified with target label, the relevant information similar to EBOM and MBOM can be presented.

The part named Movement is used for Case B. The query result suggests: All subparts belonging to Movement, the operation which produces Movement and the former operation of the queried operation, are intended to be retrieved. As presented in Fig. 3, five subparts belong to the part Movement. The operation named Movement Assembly (A1) assembles the part Movement, whereas the part Gear Set is

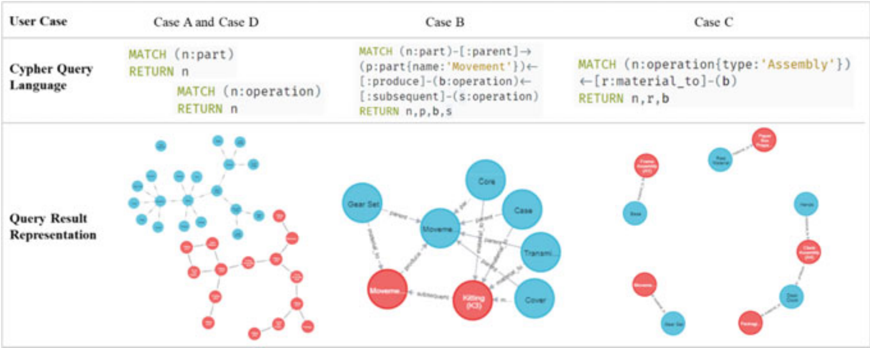


Fig. 3 Visualization of query results for use cases

direct feeding to the operation Movement Assembly (A1) and other four subparts are prepared by the former operation Kitting (K3).

The query result for Case B intends to retrieve any nodes which feed any operations belonging to the type Assembly. The results explain that there are four operations refer to the type Assembly. After the successful retrieval of all assembly operation, the required materials or parts fed to those operation are also illustrated (Fig. 3).

3.5 Discussion

The query results for the use cases above suggest that the graph model based BOM is feasible to reach the collaboration workflow illustrated in Fig. 1b. Despite the fact that the designed use cases for validation do not cover all scenarios in terms of collaboration between product development and production planning, the graph model based BOM still proves its utility and coverage of basic requirements in both departments. Characteristics and further potentials are thus discussed as follows:

Functional Fulfilment: As graph data model is based on entities and relations, it can fulfil any functions which relational data model can achieve, theoretically. However, the usage and implementation depend on specific scenarios. It would not be such competitive than relational approach, when only limited number of simple products have to be managed. Nevertheless, the utilization of graph model based BOM towards future profitability is worthy to explore such the break-even-point of using this approach.

Extendibility: A future-oriented and well-designed graph database has good extendibility due to its form of data storage. More information and knowledge can be integrated and managed by adding new nodes and relations. Possible extensions are summaries as: (a) content extension such as the storage of URL of the corresponding CAD-Model in part's property for access and editing; (b) organizational extension, for instance, a cross-departmental collaboration via graph model based BOM in terms of the needs of data interaction; (c) functional extension e.g., inference, comparison and analysis regarding BOM relevance.

Non-centric data model: Compared to traditional approaches, the graph model based BOM with numerous non-centric nodes is not built with the focus on serving any certain business unit. Instead, each business unit contributes their own information to the graph database, so that the company-wide knowledge are built. This scheme can solve the problematic which is caused by highly dependent engineering activities between up- and downstream processes.

Information continuity: Because of the character of containing both product- and production-related data and information, the graph model based BOM ensures a single source of data for cross-departmental engineering activities. Although errors caused by human and data incompleteness are not to be eliminated, the problem of information synchronization and communication problems in collaboration in engineering could be alleviated.

4 Summary and Outlook

This paper focuses on the introduction of graph model based BOM for optimized coupling of product development and production planning on an operational level. Initially, the premise is put forward that the introduction of a graph model based BOM increases the efficiency of collaboration between product development and production planning. In order to validate the concept, an exemplary BOM data is acquired, as well as adapted. In further step, a simple graph data model for storing both product and production information is designed. As result, the graph model based BOM is established utilizing the graph database Neo4j. Furthermore, four use cases as scenarios in engineering activities are defined to represent basic needs in engineering activities. Finally, the use cases are evaluated as Proof-of-Concept and the query results are presented and explained.

Considering the complexity of networked product and production data in practice, further researches should focus on applying and validating the approach based on certain industry needs. On one hand, it is necessary to answer the question, whether introducing graph model based BOM, rather than adapting and extending existing functionalities based on a traditional relational data model, could bring long-term quantifiable benefits for companies. On the other hand, it is essential to analyze and evaluate the effects of introducing graph data model of BOM for organizational process optimization, specifically the optimization of workflow at an operational level. Furthermore, a generic framework using a graph approach for BOM management can be explored concerning different products and production processes in various industries. It can enable both company-wide cross-departmental integration of BOM information, as well as enhance cross-organizational collaboration between suppliers and OEMs.

References

1. Stonebraker, P.W.: Restructuring the bill of material for productivity: a strategic evaluation of product configuration. Int. J. Prod. Econ. **45**(1–3), 251–260 (1996). https://doi.org/10.1016/0925-5273(95)00147-6
2. Hegge, H.M.H., Wortmann, J.C.: Generic bill-of-material: a new product model. Int. J. Prod. Econ. **23**(1), 117–128 (1991). https://doi.org/10.1016/0925-5273(91)90055-X
3. Kneppelt, L.R.: Product structuring considerations for master production scheduling. Prod. Invent. Manag. **25**(1), 83–99 (1984)
4. van Veen, E.A., Wortmann, J.C.: Generative bill of material processing systems. Prod. Plann. Control **3**(3), 314–326 (1992). https://doi.org/10.1080/09537289208919403
5. Chang, S.-H., Lee, W.-L., Li, R.-K.: Manufacturing bill-of-material planning. Prod. Plann. Control **8**(5), 437–450 (1997). https://doi.org/10.1080/095372897235019
6. Jiao, J., Tseng, M.M., Ma, Q., Zou, Y.: Generic bill-of-materials-and-operations for high-variety production management. Concurrent Eng. Res. Appl. **8**(4), 297–321 (2000). https://doi.org/10.1106/95P6-GB09-YHG6-H5QG
7. He, L., Ni, Y., Ming, X., Li, M., Li, X.: Integration of bill of materials with unified bill of materials model for commercial aircraft design to manufacturing. Concurrent Eng. Res. Appl. **22**(3), 206–217 (2014). https://doi.org/10.1177/1063293X14542443
8. Lee, J.H., Kim, S.H., Lee, K.: Integration of evolutional BOMs for design of ship outfitting equipment. CAD Comput. Aided Des. **44**(3), 253–273 (2012). https://doi.org/10.1016/j.cad.2011.07.009
9. He, W., Ming, X.G., Ni, Q.F., Lu, W.F., Lee, B.H.: A unified product structure management for enterprise business process integration throughout the product lifecycle. Int. J. Prod. Res. **44**(9), 1757–1776 (2006). https://doi.org/10.1080/00207540500445453
10. Xiaojun, M., Xu, Z., Ruxin, N., Aimin, W.: Task-oriented BOM conversion and view mapping. In: 14th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2010, pp. 565–568 (2010)
11. Aydin, A.O., Güngör, A.: Effective relational database approach to represent bills-of-materials. Int. J. Prod. Res. **43**(6), 1143–1170 (2005). https://doi.org/10.1080/00207540512331336528
12. Fusheng, Q., Wenjian, L.: An XML-based approach to represent generic bill of materials. In: 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, SNPD 2007, pp. 832–837 (2007)
13. Yang, Z., Zhang, J., Wang, S., Wang, J., Huang, X.: Building ontology-based bill of material design and knowledge management in power gird. In: 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016, pp. 1664–1669 (2016)
14. Chen, L., Wang, T., Song, Z.: A Web-based product structure manager to support collaborative assembly modeling. J. Comput. Inf. Sci. Eng. **4**(1), 67–77 (2004). https://doi.org/10.1115/1.1666894
15. Romanowski, C.J., Nagi, R.: A data mining and graph theoretic approach to building generic bills of materials. In: 11th Industrial Engineering Research Conference (2002)
16. Guoli, J., Daxin, G., Tsui, F.: Analysis and implementation of the BOM of a tree-type structure in MRPII. J. Mater. Process. Technol. **139**(1–3), 535–538 (2003). https://doi.org/10.1016/S0924-0136(03)00520-X
17. Kashkoush, M., ElMaraghy, H.: Matching bills of materials using tree reconciliation. Proc. CIRP **7**, 169–174 (2013). https://doi.org/10.1016/j.procir.2013.05.029
18. Wang, Y., Wang, W., Mao, S.: A new type of BOM model and its application. In: 2013 International Conference on Precision Mechanical Instruments and Measurement Technology, ICPMIMT 2013, pp. 1234–1238 (2013)
19. Yan, C., Song, Y.: Colored graph based BOM layer model. In: 2010 International Conference on Advances in Materials and Manufacturing Processes, ICAMMP 2010, pp. 1156–1160 (2010)

20. Li, Y., Tao, F., Cheng, Y., Zhang, X., Nee, A.Y.C.: Complex networks in advanced manufacturing systems. J. Manuf. Syst. **43**, 409–421 (2017). https://doi.org/10.1016/j.jmsy.2016.12.001
21. Cinelli, M., Ferraro, G., Iovanella, A., Lucci, G., Schiraldi, M.M.: A network perspective for the analysis of bill of material. In: 13th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME 2019, pp. 19–24 (2019)
22. Schmidt, M., Gehring, B., Gerber, J.-S., Stocker, J.M., Kreimeyer, M., Lienkamp, M.: Graph-based similarity analysis of BOM data to identify unnecessary inner product variance. In: 21st International Conference on Engineering Design, ICED 2017, pp. 489–498 (2017)
23. Bhattacharyya, A., Chakravarty, D.: Graph database: a survey. In: 2020 International Conference on Computer, Electrical and Communication Engineering, ICCECE 2020, pp. 1–8 (2020)
24. Guia, J., Soares, V.G., Bernardino, J.: Graph databases: Neo4j analysis. In: 19th International Conference on Enterprise Information Systems, ICEIS 2017, pp. 351–356 (2017)

Logistics Engineering and Management

Intra-Route Location Routing for the Pickup and Delivery Problem with Transfers

Cansu Agrali, Mario Ventresca, and Seokcheon Lee

Abstract The development and increasing application area of electric vehicles (EVs) in city logistics has contributed to the necessity of intermediate stops for charging EVs. Due to the limited service range of these vehicles, charging facilities must be located at the same echelon as customers' and differ from depots or hubs in that they are visited while serving customers. Thereby, these facilities allow vehicles to exchange requests and get recharged. In this study, we present the Intra-route Location Routing for the Pickup and Delivery Problem with Transfers that arises in this novel scenario. There are sets of requests, vehicles, and potential transfer locations (intra-route facilities). Each request consists of a pair of a pickup node and a delivery node. Vehicles start their routes from their respective origins, serve customers (pickup and delivery), and return to the origin nodes. In the generic Pickup and Delivery Problem, a request must be served by a single vehicle that picks and delivers the request. In our study, however, multiple vehicles can service a request collaboratively by transferring it to each other at a transfer facility. In this problem context, we decide vehicle schedules, transfer decisions, and which nodes should serve as intra-route facilities. We propose a mathematical model for this novel NP-Hard problem. Experimental results indicate the computational difficulty of the problem in practice. For several small instances, solving the mathematical model by using a commercial solver does not even find the optimal solutions in six hours, and given the importance of this problem, it must be addressed in future works.

Keywords Location routing problem · Transfers · Pickup and delivery · Mixed integer linear programming

C. Agrali (✉) · M. Ventresca · S. Lee
Purdue University, West Lafayette, IN, USA
e-mail: cagrali@purdue.edu

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_56

1 Introduction

Logistic companies have been facing difficulties to provide low-cost delivery service with the increase of online sales [1]. Crowdsourced delivery, defined as "where ordinary people carry out last-mile deliveries with their own vehicles, from stores or warehouses to customers' destinations" [2], is one plausible approach to improve such systems. Consequently, several companies already launched their own crowdsourced delivery services [3, 4].

A negative outcome of the increased sales and deploying more vehicles, however, is the environmental impact [5]. Electric vehicles (EVs) play an important role to curtail the carbon emission from the conventional vehicles [6]. The study in [7] reveals that EVs do not only have a long-term environmental benefit, but also reduce the operational cost by 39–60% in the short term. However, EVs have several challenges including a short service range [8] and thus, they must be recharged during their daily tasks to stay operational. The facilities equipped with such charging stations called as *intra-route facilities* in the literature.

In this paper, we study a novel *Pickup and Delivery Problem* (PDP). There is a set of requests to be served by picking them up from their origin nodes and delivering them to their destinations. Note that, a request is associated with a pair of particular pickup and delivery nodes. Therefore, random picked up request cannot be used to satisfy the demand in a specific delivery node. Pickup and delivery locations have their available time windows (PDP-TW). A fleet of EVs is used for the transportation services. Each vehicle belongs to a particular depot, and there may be multiple depots. In the generic PDP, a request is picked up and delivered by a single vehicle; however, exchanging requests at an intra-route facility is allowed in this study, which is called *transfers* in the literature (PDP-T).

Figure 1 presents an illustrative example. There are two vehicles' routes (solid and dashed lines) and three requests (r_1, r_2, r_3). The circles with labels p and d represent pickup and delivery nodes, respectively. Together p_1 and d_1 define one request (r_1). The black triangle demonstrates the intra-route facility and the square is for the depot; r_1 and r_2 are served by a single vehicle, and r_3 is transferred between vehicles at the intra-route facility. One vehicle picks up the request r_3 and brings it to the facility, and the other vehicle transports r_3 from the facility to its destination.

1.1 Relevant Literature

With the increasing utilization of EVs, the *Vehicle Routing Problem with Intermediate Stops* (VRP-IS) has been increasingly studied. The intermediate stops refer to intra-route facilities in our study. The VRP-IS was first studied by the authors of [9, 10]; an *Adaptive Large Neighborhood Search* algorithm was developed in [9], and the first *Mixed Integer Linear Programming* (MILP) model was proposed in [10]. In these studies, the locations of intra-route facilities are already known, and they analyzed the

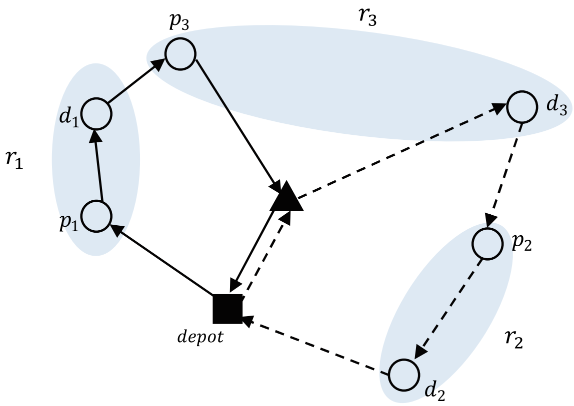


Fig. 1 Example solution for the PDP-T: r_1 and r_2 are not transferred, while r_3 is transferred between vehicles

economic and environmental impact of EVs. The authors of the study [10] presented an ALNS algorithm for the *Location Routing Problem with Intra-route Facilities* (LRP-IF) in [11]. They claimed that they find the new best known solutions to the problem. In these recent studies, the VRP model was applied, and the model differs from the PDP in the request descriptions. Specifically, a request in the PDP model is associated with a pickup and a delivery node, where the pickup node must be visited before the associated delivery node. The LRP with PDP was first discussed in [12] and the authors proposed a *Column Generation* model to solve it and provided several sensitivity analyses on the fixed and route costs.

1.2 Our Contribution

To the best of our knowledge, we are the first to study the *Location Routing Problem with Intra-route Facilities, Transfers, and Pickup and Deliveries* (LRP-PDP-IF-T). Our contribution is considering intra-route facilities and transfers in the LRP-PDP. Our problem also covers multiple depots, partial charging, and a heterogenous fleet. In Sect. 2, the problem definition and the MILP model are presented, and computational experiments are shown in Sect. 3.

2 Problem Definition and Mathematical Model

The problem is defined on a network that includes sets of nodes: depots, candidate transfer nodes, and request nodes. Each request ($r \in R$) has several attributes including a pickup node—$p(r)$ and a delivery node—$d(r)$. Vehicles' origin and

Table 1 Sets

N:	Set of all nodes
D:	Set of depots (includes final depots)—$D \subset N$
T:	Set of candidate transfer nodes—$T \subset N$
R:	Set of requests
V:	Set of vehicles

Table 2 Parameters

c_i:	The cost of locating a transfer at node $i \in T$
t_{ij}:	The distance between node $i \in N$ and node $j \in N$
Q_k:	The capacity of vehicle $k \in V$
B_k:	The maximum mileage that vehicle $k \in V$ travels
H_k:	The mileage that vehicle $k \in V$ can travel with 1-min charging
L_i:	The load of node $i \in N$
E_i:	The earliest time that a vehicle can arrive to node $i \in N$
T_i:	The latest time that a vehicle can arrive to node $i \in N$
M:	A big number

destination depots are indicated by $o(k)$ and $o'(k)$, respectively. They are the exact same locations. Vehicle $k \in V$ starts its route from $o(k)$, and end at $o'(k)$. Other sets and parameters are presented in Tables 1 and 2. The decision variables are indicated in Table 3.

Minimize

$$\sum_{i \in N} \sum_{j \in N: i \neq j} \sum_{k \in V} t_{ij} x_{ij}^k + \sum_{i \in T} c_i w_i \tag{1}$$

Subject to

Table 3 Decision variables

x_{ij}^k:	Binary variable, gets 1 if vehicle $k \in V$ directly travels from node $i \in N$ to node $j \in N$
y_{ij}^{kr}:	Binary variable, gets 1 if request $r \in R$ is carried by vehicle $k \in V$ from node $i \in N$ to node $j \in N$
z_{ir}^{kl}:	Binary variable, gets 1 if request $r \in R$ is transferred from vehicle $k \in V$ to $l \in V$ at transfer node $i \in T$
w_i:	Binary variable, gets 1 if an intra-route facility is located at transfer node $i \in T$
p_i^k:	Vehicle $k \in V$'s charging duration at transfer node $i \in T$
a_i^k:	Arrival time of vehicle $k \in V$ to node $i \in N$
d_i^k:	Departure time of vehicle $k \in V$ from node $i \in N$
b_i^k:	Remaining battery of vehicle $k \in V$ when it arrives at node $i \in N$

$$\sum_{j \in N} x_{ij}^{k} = 1 \quad \forall k \in V, i = o(k) \tag{2}$$

$$\sum_{j \in N} x_{ij}^{k} = \sum_{j \in N} x_{jl}^{k} \quad \forall k \in V, i = o(k), l = o'(k) \tag{3}$$

$$\sum_{j \in N} x_{ij}^{k} = \sum_{j \in N} x_{ji}^{k} \quad \forall i \in N \backslash \{o(k), o'(k)\}, k \in V \tag{4}$$

$$\sum_{j \in N} \sum_{k \in V} y_{ij}^{kr} = 1 \quad \forall r \in R, i = p(r) \tag{5}$$

$$\sum_{j \in N} \sum_{k \in V} y_{ij}^{kr} = 1 \quad \forall r \in R, i = d(r) \tag{6}$$

$$\sum_{j \in N} y_{ij}^{kr} = \sum_{j \in N} y_{ji}^{kr} \quad \forall i \in N \backslash \{p(r), d(r), T\}, k \in V, r \in R \tag{7}$$

$$\sum_{j \in N} \sum_{k \in V} y_{ij}^{kr} = \sum_{j \in N} \sum_{k \in V} y_{ji}^{kr} \quad \forall i \in T, r \in R \tag{8}$$

$$x_{ij}^{k} \le w_j \quad \forall i \in N, j \in T, k \in V : i \ne j \tag{9}$$

$$y_{ij}^{kr} \le x_{ij}^{k} \quad \forall i, j \in N, k \in V, r \in R : i \ne j \tag{10}$$

$$d_i^k + t_{ij} - a_j^k \le M\left(1 - x_{ij}^k\right) \quad \forall i, j \in N, k \in V : i \ne j \tag{11}$$

$$a_i^k \le d_i^k \quad \forall i \in N \backslash D, k \in V \tag{12}$$

$$a_i^k + p_i^k \le d_i^k \quad \forall i \in T, k \in V \tag{13}$$

$$d_i^k \ge E_i \quad \forall k \in V, r \in R : i = p(r) \tag{14}$$

$$a_i^k \le T_i \quad \forall k \in V, r \in R : i = d(r) \tag{15}$$

$$\sum_{j \in N} \left(y_{ji}^{kr} - y_{ij}^{kr}\right) + \sum_{j \in N} \left(y_{ij}^{lr} - y_{ij}^{lr}\right) \le 1 + z_{ir}^{kl} \quad \forall i \in T, k, l \in V, r \in R : k \ne l \tag{16}$$

$$a_i^k - d_i^l \le M\left(1 - z_{ir}^{kl}\right) \quad \forall i \in T, k, l \in V, r \in R : k \ne l \tag{17}$$

$$z_{ir}^{kl} \le w_i \quad \forall i \in T, k, l \in V, r \in R : k \ne l \tag{18}$$

$$\sum_{r \in R} y_{ij}^{kr} \leq Q_k x_{ij}^k \quad \forall i, j \in N, k \in V : i \neq j \tag{19}$$

$$b_j^k \leq b_i^k - t_{ij} + M\left(1 - x_{ij}^k\right) \quad \forall i \in N \backslash T, j \in N, k \in V : i \neq j \tag{20}$$

$$b_j^k \leq b_i^k - t_{ij} + M\left(1 - x_{ij}^k\right) + H_k p_i^k \quad \forall i \in T, j \in N, k \in V : i \neq j \tag{21}$$

$$b_i^k = B_k \quad \forall k \in V, i = o(k) \tag{22}$$

$$b_i^k + H_k p_i^k \leq B_k \quad \forall i \in T, k \in V \tag{23}$$

$$p_i^k \leq (B_k / H_k) \sum_{j \in N} x_{ji}^k \quad \forall i \in T, k \in V \tag{24}$$

$$x_{ij}^k, y_{ij}^{kr}, z_{tr}^{kl} \in \{0, 1\} \quad \forall i, j \in N, t \in T, k, l \in V, r \in R : i \neq j, k \neq l \tag{25}$$

$$a_i^k, d_i^k, b_i^k, p_t^k \geq 0 \quad \forall i \in N, t \in T, k \in V \tag{26}$$

The objective is minimizing the total distance travelled and the cost of locating intra-route facilities (1). Constraints (2) ensure that all vehicles leave their depot, and each vehicle has only one route. All vehicles must return to their depot in (3). Although two separate nodes are used for the origin and the destination depots, in the network, they are located at the same coordinates. Vehicles can directly travel from their origin depot to the destination, which would incur zero cost to the system. Constraints (4) capture the flow balance in all nodes except for the vehicle's origin and the destination depots. All requests must be picked up and delivered respectively in Constraints (5) and (6). Constraints (7) ensure that if a request visits a node that is not its pickup, its delivery, or a transfer node, then the request must leave that node with the same vehicle. On the other hand, if a request visits a transfer node, it must leave the transfer node weather with the same vehicle or with another vehicle in (8). If a facility is not opened at a transfer node, the node cannot be visited in (9). Constraints (10) capture the relationship between two variables, in order to carry a request from node i to node j, the vehicle must travel through the arc (i, j). The arrival and departure times are calculated in Constraints (11). Vehicles cannot leave a node earlier than their arrival time (12), but if they are recharged at a transfer node, the departure time must be greater than or equal to the arrival time plus the recharging duration at the transfer node (13). The time window constraints are captured in (14) and (15). Constraints (16) ensure that if a request is brought to a transfer node by a vehicle (k) and leaves the node with another vehicle (l), then the corresponding z variable becomes one. Constraints (17) capture the arrival and departure times of vehicles, exchanging requests, at transfer nodes. If an intra-route facility is not located in a transfer node, then exchanging requests is not possible (18). Vehicle capacity constraints are captured by (19). Battery consumptions are calculated after a regular and a transfer node by (20) and (21), respectively. Vehicles leave their origin depot fully charged (22). Constraints (23) ensure that vehicles cannot be recharged more than their battery capacity, and vehicles can only be recharged if they visit an intra-route facility (24). Constraints (25) capture binary variables while (26) declares positive continues variables.

3 Computational Study

We generated different-sized instances to run the mathematical model and evaluate the problem complexity. The smallest instances contain five requests, one depot, and two potential transfer locations. The model does not open any transfer facilities in such small networks as vehicles can serve to all requests without recharging. For larger instances of 10 or more requests, the model does not find a feasible solution in three hours. Therefore, we modified the smallest instances by doubling the distances between nodes so that the model may decide to open transfer facilities to charge vehicles and/or to transfer requests. The data set can be accessed through this link. The tests are performed on a PC equipped with Intel(R) Core(TM) i7-4790 CPU running at 3.60 GHz, with 16 GB RAM.

Preliminary results suggest that increasing distances has a significant impact on the computational run-time: the model cannot find the optimal solution in 24 h. In order to better understand the problem difficulty, we solve the model with a six-hour time limit. With this, the quality of the solutions, obtained after six hours, can be discussed. The results are presented in Table 4. Three sets of instances are used. ‘e’, ‘m’, and ‘h’ represent easy, medium, and hard, respectively. All instances consist of five requests and one depot; however, easy and medium instances have four potential transfer locations, while hard instances have only two. Instances are labeled based on the distances between nodes and load of requests. The closer the distances are, the easier the instance is as EVs have battery limitation, and nodes have time-window constraints. Five random instances are generated for each set. In the table, the columns demonstrate the objective function value, the solution time, the percentage relative gap (the gap between the upper bound and the lower bound when the model stops), the total number of facilities opened, the total number of requests transferred, and the number of vehicles deployed, respectively.

The MILP model finds the optimal solution in six of 15 instances in six hours. However, any feasible solutions cannot be found for four instances: h2–h5. Additionally, although at least one intra-route facility is opened in nine instances, they are not used to transfer requests. The results highlight the importance of intra-route facilities to recharge EVs even for small networks.

Table 4 Solutions from MILP

Instance	Objective value	Time (s)	%*Gap*	Facilities opened	Requests transferred	Vehicles deployed
e1	227.22	2925	0	0	0	1
e2	276.90	642	0	1	0	1
e3	234.48	2492	0	1	0	1
e4	296.00	21,187	0	1	0	1
e5	311.63	*TL*	20.00	1	0	1
m1	328.96	10,154	0	1	0	1
m2	274.38	446	0	0	0	1
m3	477.01	*TL*	18.04	2	0	1
m4	287.16	*TL*	7.12	1	0	1
m5	425.37	*TL*	34.73	2	0	1
h1	653.04	*TL*	13.37	1	0	2
h2	–	–	–	–	–	–
h3	–	–	–	–	–	–
h4	–	–	–	–	–	–
h5	–	–	–	–	–	–

* *TL* refers to the time limit, which is six hours, 21,600 s

4 Conclusion

To the best of our knowledge, we are the first to study the LRP with transfers, EVs, and pickup and delivery. We propose a MILP model for the problem. The importance of the problem is discussed in Sect. 1. Even though solving the problem to optimality is crucial, the computational experiments in this paper highlight the problem difficulty in practice as even the small instances in this paper were difficult to solve. Our follow-up research is developing metaheuristic approaches to solve the problem more efficiently. Further research could explore valid inequalities to improve the runtimes and the objective values of the MILP. Moreover, the problem can potentially be extended by including drone deliveries since drones also have battery and capacity constraints but stricter.

References

1. Arslan, A.M., et al.: Crowdsourced delivery—a dynamic pickup and delivery problem with ad hoc drivers. Transp. Sci. **53**(1), 222–235 (2019)
2. Alnaggar, A., Gzara, F., Bookbinder, J.H.: Crowdsourced delivery: a review of platforms and academic literature. Omega **98**, 102139 (2021)
3. Halzack, S.: Amazon Flex: the retailer's Uber-like effort to bring you packages. Washington Post. Retrieved 1 April 2021. https://www.washingtonpost.com/news/business/wp/2015/09/29/amazon-flex-the-retailers-uber-like-effort-to-bring-you-packages/ (2015).

4. Bose, N.: Walmart trials grocery delivery to rival Amazon Flex. Reuters. Retrieved 1 April 2021. https://www.reuters.com/article/us-walmart-delivery/walmart-trials-grocery-delivery-to-rival-amazon-flex-idUSKCN1LL1Q1 (2018)
5. Bellos, I., Ferguson, M., Beril Toktay, L.: The car sharing economy: interaction of business model choice and product line design. Manuf. Serv. Oper. Manage. **19**(2), 185–201 (2017)
6. Chen, T.D., Kockelman, K.M., Hanna, J.P.: Operations of a shared, autonomous, electric vehicle fleet: implications of vehicle and charging infrastructure decisions. Transp. Res. Part A Policy Pract. **94**, 243–254 (2016)
7. Sahin, B., et al.: An approach for analysing transportation costs and a case study. Eur. J. Oper. Res. **193**(1), 1–11 (2009)
8. Lin, J., Zhou, W., Wolfson, O.: Electric vehicle routing problem. Transp. Res. Proc. **12**, 508–521 (2016)
9. Yang, J., Sun, H.: Battery swap station location-routing problem with capacitated electric vehicles. Comput. Oper. Res. **55**, 217–232 (2015)
10. Schiffer, M., Walther, G.: The electric location routing problem with time windows and partial recharging. Eur. J. Oper. Res. **260**(3), 995–1013 (2017)
11. Schiffer, M., Walther, G.: An adaptive large neighborhood search for the location-routing problem with intra-route facilities. Transp. Sci. **52**(2), 331–352 (2018)
12. Capelle, T., et al.: A column generation approach for location-routing problems with pickup and delivery. Eur. J. Oper. Res. **272**(1), 121–131 (2019)

Method for Determining Loading Positions of Delivered Parcels by Genetic Algorithm with Three-Dimensional Modified BL Method with Multidirectional Reference Points

Eiji Harayama, Yoshinari Yanagawa, and Ikuo Arizono

Abstract In recent years, the number of delivered parcels has rapidly increased along with the growth of the Internet and courier services have been booming. In addition, many people refrained from going out due to the spread of the COVID-19, and also people who have never used e-commerce have started using e-commerce. Therefore, it has been become to require to deliver parcels more efficiently. One of the evaluation measures to realize an efficient delivery is minimizing the burden of loading/unloading parcels. To do so, loaded positions of parcels should be decided by considering the unloading sequence of the parcels so that a delivery person moves few parcels at the time of unloading them. Hence, some systematic method for deciding loaded positions of parcels would be required. However, at present, the delivery person determines the loading positions of them with not a systematic method but his experience or intuition method. In this research, we would try to optimize loading positions of parcels by using genetic algorithm. Our proposed method of determining the loading position would use genetic algorithm with three-dimensional modified bottom-left method with multidirectional reference points. In addition, we would make comparison between the method of loading from only one direction and the method of loading from multiple directions. The number of planned movement of parcels would be treated as an evaluation measure as well as previous researches. We would make comparison between the precedent and the proposed method, and consideration of the results. We could conclude that our proposed method has dramatically improved the precedent methods. The results by the multiple directions were about 20% better than the ones by one direction.

Keywords Genetic algorithm · Packing problem · Courier service

E. Harayama (✉) · Y. Yanagawa · I. Arizono
Okayama University, Okayama, Japan
e-mail: put670l1@s.okayama-u.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_57

1 Introduction

In the 1970s, courier services in current form are launched. Since then, courier services diversify and expand to meet a variety of demand, and they have achieved rapid growth for decades. In the last five years up to 2020, the volume of courier services in Japan increases by about 15.4%, and courier services are expected to continue to expand in the future. One of the reasons for this rapid growth is that a lot of not only major companies but small business operators make inroads in the electronic commerce market. In addition, flea market applications develop rapidly on the spread of the Internet and smartphones, and the growth of electronic commerce is still continuing presently. In addition, because the Japanese government asks people to quarantine themselves from April 2020 due to the spread of the COVID-19, the online shopping is expanded to consumers who don't previously use it. As a result, the demand for courier services increases rapidly. It is expected that the expansion of courier services continues even after the COVID-19 settles down. Consequently, the demand of delivery process efficiency is increasing in the courier service.

In the delivery process, delivered parcels are transported to its destination by truck or container. The transportation of delivered parcels is divided into two stages. The first stage is to distribute and transport delivered parcels to each sales office, and it is required for this process to transport a large number of delivered parcels in a shipment. In other words, the loading result is evaluated by the volume utilization, which is the percentage of the total parcels' volume in a truckload capacity. Various loading methods have already been studied to maximize the volume utilization [1–3]. The second stage is to transport delivered parcels from each sales office to delivery destinations, and it is necessary to consider not the volume utilization but the order in which delivered parcels are unloaded. Currently, this consideration is based on the experience and intuition of each workers. Therefore, they cannot determine the optimal loading position of delivered parcels. As a result, they may need to reload some parcels that are encumbering the parcel which should be unloaded. These delivery operations are inefficient.

The problem defined in this study is classified as a kind of packing problems. In packing problems, several objects must be placed in a limited area under various constraints. The problems are distinguished into two types: two-dimensional packing problems such as rectangular packing or polygonal packing, and three-dimensional packing problems such as rectangles and spheres. These problems are one of combinatorial optimization problems and have been studied extensively. For two-dimensional packing problem, various algorithms have been constructed that are able to be applied to an enormous range of solutions. The two-dimensional packing problems are applied to the design of integrated circuits or the materials industry. The three-dimensional packing problems are more complex than the two-dimensional packing problems because the number of dimensions increases. A lot of studies for maximizing the volume of objects for a limited volume have been studied, and various methods have been proposed to obtain approximate solutions for the problems of practical scale. Typical examples of such methods are Deepest-Bottom-Left method (DBL method) or best-fit method. On the other hand, few studies have taken the order of unloading into account. The purpose of this study is to construct an algorithm for determining loading positions of delivered parcels taking the order of unloading into account.

2 Problem Formulation

2.1 Scale Setting

The loading space for delivered parcels and the size of them were set based on the type of vehicle and the size and number of delivered parcels used in actual courier services. The shapes of delivered parcels were rectangles or cubes and delivered parcels were considered homogeneous.

The loading space for delivered parcels had a width of 130 cm, a depth of 180 cm, and a height of 120 cm. The space was embedded in a three-dimensional Cartesian coordinate system. The coordinate origin in the first octant of the system was treated as the origin point in the loading space. The width, the depth and the height of the space were oriented in accordance with the directions of the x-axis, the y-axis, and z-axis. Delivered parcels were taken out in the rear direction of the space, that is, in the positive direction of the y-axis.

The size of each delivered parcels was within 60 cm in width, in depth, and in height. In addition, the total length of the three sides was within 60 cm to 150 cm, and they were randomly generated in increments of 10 cm. Further, the depth of delivered parcels was longer than the width of them. This is because it had been known from previous studies that setting the depth of them longer than the width is more efficient. The total number of delivered parcels was 40, referring to the actual courier services.

2.2 Restrictions on Loading Delivered Parcels

The following constraints (1)–(4) are set based on the actual courier services.

(1) Placement constraints
Each of delivered parcels have to be loaded inside the loading space for delivered parcels.
(2) Parallel constraints
Delivered parcels must be loaded parallel to the x-axis, y-axis, and z-axis of the loading space for delivered parcels, that is, they cannot be loaded diagonally with respect to the space.

(3) Orientation constraints
In courier services, in order to prevent damage to delivered parcels, it is common to avoid rotation of delivered parcels, such as changing the height or reversing the top and bottom. Hence, no such rotation is performed.
(4) Stability constraints
Considering the vibrations caused by driving during delivery operations, in the case that delivered parcels are stacked on top of other, those are stacked in a stable condition. The stability of delivered parcels is evaluated as the ratio of the bottom area in contact with below them to the total bottom area of them. In this study, assume that delivered parcels can be stacked if and only if the ratio is greater than or equal to 0.7, due to the previous study [4].

2.3 Evaluation of Loading Results

The previous study [5] has used the total number of expected transshipment parcels as an evaluation measure. Similarly, in this study, the total number of expected transshipment parcels employs the evaluation measure. The following is the description of expected transshipment parcels.

When delivery workers unload one delivered parcel at delivery customer's location, they must move the others that are involved with the delivered parcel. Such moved parcels that are obstructions are considered as expected transshipment parcels. In this study, the loading result is considered better as the total number of expected transshipment parcels is smaller. The following (1)–(3) shows the conditions of subject to expected transshipment parcel.

(1) A parcel that is located above an unloading parcel and that obstructs the unloading.
(2) A parcel that is located in the direction of removal from an unloading parcel and that obstructs the unloading.
(3) A parcel that indirectly obstructs an unloading parcel.

The loading position of delivered parcels may change during a delivery, and the total number of transshipment parcels may change accordingly. In this study, we do not take the change of the loading position during operations in the delivery into account. Therefore, we calculate the total number of expected transshipment parcels for the loading position at the delivery departure. When the number is small, the loading position quite maintains the order of unloading.

3 The Method of the Previous Studies

In the previous study [5], vertex management method and Deepest-Bottom-Left method (DBL method) have been used as the basic loading algorithms.

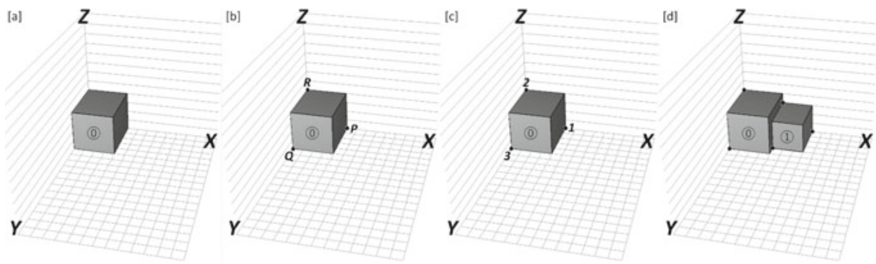


Fig. 1 Procedure of loading delivered parcels

In vertex management method, three reference points have been generated from the vertices of the already loaded parcels [6]. As shown by the circles in Fig. 1b, three points have been generated from the already loaded parcels in the X, Y, and Z axis directions. By choosing one of these points, the position of the next parcels has been determined.

DBL method has been an extension of the BL method to three dimensional and has loaded parcels as close as possible to the bottom and left (in that order) of the deepest part. In this method, the loading position of a parcel has been determined by selecting one of the reference points generated above.

Delivered parcels are loaded in the reverse order of unloading. That is, delivered parcels should be loaded in order of decreasing delivered parcels order. The procedure of loading parcels in the conventional method shows as follows:

(1) Place parcel 0 at the origin (Fig. 1a).
(2) Generate three reference points from the loaded parcel according to vertex management method (Fig. 1b).
(3) Order the reference points to be loaded according to DBL Method (Fig. 1c).
(4) Determine whether the next parcel is loaded or not in the order specified in (3). At the point where the restrictions (2.2) are satisfied, load the reference point with the parcel. Reference points that are loaded with the parcel are deleted (Fig. 1d). If the restrictions (2.2) are not satisfied for all reference points, then it is determined that loading is not allowed.
(5) Repeat steps (2)–(4) until all the delivered parcels are loaded.

In the previous method [5], a different algorithm (EVR) has been applied in addition to DBL method to reduce the total number of expected transshipment parcels. When packing one parcel by vertex management method and DBL method, each reference point has been searched to see if there are any expected transshipment parcels in the direction of outlet when the parcel is tentatively placed. After that, the loading position of the parcel have been determined according to the existence of expected transshipment parcels and the number of obstructing parcels. Specifically, the weighted value of reference points by DBL method has been multiplied by the number of expected transshipment parcels. EVR has decided the loading position of a parcel in accordance with the multiplied weighted value of reference points in ascending order.

4 Proposal Method

4.1 Three-Dimensional Modified BL Method with Multidirectional Reference Points

In the proposed method, different Three-dimensional modified BL method has been used in addition to DBL method. A total of twelve methods have been used: six in which the priority order of D, B, and L are swapped, and another six in which L (Left) are replaced by R (Right). For example, DLB Method loads parcels as close as possible to the left and bottom (in that order) of the deepest part. In addition, DBR Method swaps left and right, and loads parcels as close as possible to the bottom and right (in that order) of the deepest part.

4.2 Adaptation of Genetic Algorithm

We have genetically engineered twelve different methods (4.1) to apply genetic algorithm (GA) and generated the gene sequence by assigning a gene from 0 to 11 to each parcel. Each parcel has been loaded using a different method according to its gene. Specifically, DBL, DLB, BDL, LDB, BLD, LBD, DBR, DRB, BDR, RDB, BRD, and RBD method have been applied in order from 0 to 11. Table 1 shows an example of one gene sequence. In the sequence, parcel 0, 1, and 2 have been loaded using BDR method, DBL method, and RBD method. In the proposed method, we have generated an arbitrary number of gene sequences for each individual and optimized them by genetic manipulation (4.3).

4.3 Genetic Manipulation

In this chapter, we showed the details of genetic manipulation.

Selection, a form of genetic manipulation, creates a population of the next generation based on the fitness of individuals to produce a new generation. We have used roulette selection as the method of selection. Roulette selection is a method in which the probability of selection increases in proportion to the degree of adaptation of an individual. In this study, when the total number of individuals and the evaluation value of individual i were set as M and W_i, the probability P_A that is the selection

Table 1 An example of a gene sequence

Parcel number	0	1	2	· ·	37	38	39
Gene	8	0	11	· ·	7	5	2

Table 2 An example of crossover

Parcels number	0	1	2	3	4	5	6	7	8	9
Parent A	8	0	11	10	3	1	10	2	4	2
Parent B	3	4	1	10	9	8	3	1	0	2
Offspring A	8	0	11	10	9	8	3	1	0	2
Offspring B	3	4	1	10	3	1	10	2	4	2

Table 3 An example of mutation

Parcels number	0	1	2	3	4	5	6	7	8	9
A gene before mutation	8	0	11	10	9	8	3	1	0	2
A gene after mutation	8	0	11	10	9	8	3	9	0	2

probability of individual A has been expressed as

$$P_A = \frac{1}{W_A} / \sum_{i=1}^{M} \frac{1}{W_i} \tag{1}$$

The inverse of the evaluation value (the total number of expected transshipment parcels) has been used to select individuals, because an individual with smaller evaluation value should have larger selection probability. The evaluation value for individuals that fail to load has set to 99, leaving a small probability of selecting an undesirable individual.

Crossover is an operation that crosses the genes of two parents and is performed after two individuals are selected. In this study, we have used the single point crossing. We have randomly selected a cut and have exchanged pairs around the cut in pairs that cross each other. In the example of Table 2, the point between Parcel 3 and Parcel 4 has been selected as a cut, and the offspring A, B have been produced by Crossover.

Mutation is an operation to add diversity to avoid premature convergence in finding the optimal solution. In this study, for every newly generated individual, this operation has been performed with a defined probability. The mutation position has been randomly determined and its gene has been changed. In the example of Table 3, Parcel 7 has been selected as the mutation position, and its gene has been randomly changed.

5 Result and Discussion

We have simulated the results using the conventional and proposed methods with examples of 20 delivered parcels groups used in the previous study [5]. We also have compared GA-12genes with three-dimensional modified BL method with both L (left) and R (right) elements to GA-6genes only with L (left) elements. The parameters required for genetic manipulation have been set as Population: 30, Generations: 5000, Probability of Crossover: 0.70, Probability of Mutation: 0.05.

Table 4 shows the loading ratio of the delivered parcels groups and the evaluation values for previous method (EVR) and proposal method. In the proposed method, the mean values of 20 simulations in a delivered parcel group example have been shown. Unlike EVR, GA have been able to load all parcels for all delivered parcels groups. Both GA-6genes and GA-12genes have been superior to EVR in 10 out of 20 examples comparing for the average. GA-12genes have been superior to GA-6genes in 19 examples. GA is probably more practical than EVR in that GA is able to be used for a variety of delivered parcels groups. As a result, in other simulations of delivered parcels groups, GA-12genes have been able to load with loading ratio of less than 74%. Table 5 shows the evaluation values for EVR and GA-12genes. GA-12genes have been superior to EVR in 16 out of 20 examples comparing for the minimum. On the other hand, GA-12genes is not always an efficient algorithm, as variations are observed in each trial. The mean values of the computation time in EVR, GA-6genes, and GA-12genes are about 11.3[s], 396.0[s], and 1493.3[s], respectively.

Table 4 The result for each method

Delivered parcels group	Loading ratio	EVR	GA-6genes	GA-12genes
1	0.45	14	9.7	3.4
2	0.46	14	6.5	3.7
3	0.49	N/A	11.8	9.6
4	0.50	8	5.6	7.3
5	0.53	1	11.3	9.2
6	0.53	N/A	14.0	11.1
7	0.53	5	13.2	7.3
8	0.53	N/A	10.5	7.0
9	0.54	6	15.1	12.0
10	0.54	N/A	8.6	7.4
11	0.54	7	17.9	12.3
12	0.55	0	15.1	11.2
13	0.55	1	12.0	9.0
14	0.58	N/A	10.8	7.2
15	0.59	13	26.4	20.9
16	0.60	4	22.7	21.2
17	0.60	N/A	22.8	17.0
18	0.61	3	30.6	28.0
19	0.63	11	35.5	30.0
20	0.72	N/A	40.8	31.9

Table 5 The result for EVR and GA-12genes

Delivered parcels group	Loading ratio	EVR	Minimum	Mean	Maximum	SD
1	0.45	14	1	3.4	9	1.85
2	0.46	14	0	3.7	11	2.67
3	0.49	N/A	2	9.6	22	4.39
4	0.50	8	0	7.3	16	3.94
5	0.53	1	0	9.2	17	4.95
6	0.53	N/A	2	11.1	26	6.28
7	0.53	5	2	7.3	16	4.01
8	0.53	N/A	2	7.0	12	3.02
9	0.54	6	4	12.0	20	4.93
10	0.54	N/A	1	7.4	17	4.10
11	0.54	7	5	12.3	29	6.78
12	0.55	0	1	11.2	21	4.69
13	0.55	1	1	9.0	16	3.90
14	0.58	N/A	2	7.2	18	4.21
15	0.59	13	5	20.9	42	10.77
16	0.60	4	8	21.2	38	9.45
17	0.60	N/A	3	17.0	33	6.76
18	0.61	3	12	28.0	50	10.72
19	0.63	11	10	30.0	58	12.65
20	0.72	N/A	11	31.9	59	11.43

6 Conclusions

In the previous studies on packing problems, most of them have focused on maximizing volume utilization. In this study, we have constructed a loading algorithm that considers the order of unloading without taking maximizing volume utilization into account. We have used twelve Three-dimensional modified BL method with multi-directional reference points in addition to DBL method and have applied genetic algorithms. As a result, we have been able to reduce the total number of expected transshipment parcels in almost certain parcels groups. The time to calculate the loading positions has been short enough to be practical for real courier services. On the other hand, the loading results vary in each case because of the randomly generated default value on the population. In addition, the variation of the evaluation values is larger as the loading ration is higher. The future work is to construct an algorithm that can realize small variation for each trial and cope with high loading ratio.

References

1. Hermann, G., Andreas, B.: A parrallel genetic algorithm for solving the container loading problem. Int. Trans. Oper. Res. **9**, 497–511 (2002). https://doi.org/10.1111/1475-3995.00369
2. Shinji, I., Yannanhu, H., Itsuki, H.: Introduction to shape packing algorithms in python. Oper. Res. Manage. Sci. Commun. Oper. Res. Soc. Jpn. **1**, 762–769 (2002) (in Japanese)
3. Andreas, B.: A genetic algorithm for solving the container loading problem. Eur. J. Oper. Res. 814–837 (2006). https://doi.org/10.1111/J.1475-3995.1997.TB00095.X
4. Yasunori, E., Koki, H., Yukihiro, H.: Metaheuristic algorithms for container loading problem by grouping objects. Jpn. Soc. Fuzzy Theory Intell. Inform. 859–866 (2007). https://doi.org/10.3156/jsoft.18.859 (in Japanese)
5. Nobuhiro, S., Yoshinari, Y., Ikuo, A.: A method for determining the stacking position of courier parcels for terminal transport considering the pickup order and outlet. Jpn. Ind. Manage. Assoc. 190–191 (2016) (in Japanese)
6. Koki, H., Yukihiro, H., Sadaaki, M., Yasunori, E.: Algorithms for container loading problem using grouping objects. In: 22nd Fuzzy System Symposium, pp. 927–930 (2006). https://doi.org/10.14864/fss.22.0.228.0 (in Japanese)

Redesigning the Current Inventory Management Process for an SME

Cecilia Montt, Paloma Lillo, Luis Quezada, Astrid Oddershede, and Alejandra Valencia

Abstract Companies have different types of processes in their logistics chain. Thus, the added value delivered to each link generates a fundamental competitive tool, such as excellence and differentiation in the delivery service, coupling with technology and anticipating what customers need or expect, which means that companies must have an efficient inventory and resource management. On the other hand, companies need to procure goods and services for the development of their activities. These supplies are accumulated in the companies and must be managed for their correct handling and conservation. The problem that the SME has is that when customers buy in the store, in several cases, there is no product so sometimes the sale is lost, or they wait and return the day that the product is in the store. So in this work we study the case of an SME, companies that need inventory management to avoid losing customers. In this case, the inventory and sales process is modeled and simulated using Bizagi Modeler, where it was possible to obtain the breaks. Demand forecasting was studied with neural networks. Since the behavior of the demand is variable during the year, mainly December, February, Mother's Day and others, which causes peaks in demand, to analyze this behavior we used the Periodic Review P Model, where we obtained the quantity to be requested to the supplier in different periods of the year, which increases sales, since 100% of the demand is covered. The results of the simulation with BIZAGE with the current and proposed situation show that in the

C. Montt · P. Lillo · L. Quezada (✉) · A. Oddershede
Universidad de Santiago de Chile, Santiago, Chile
e-mail: luis.quezada@usach.cl

C. Montt
e-mail: cecilia.montt@usach.cl

P. Lillo
e-mail: paloma.lillo@usach.cl

A. Oddershede
e-mail: astrid.oddershede@usach.cl

A. Valencia
Universidad Catolica de Valparaiso, Valparaiso, Chile
e-mail: alejandra.valencia@pucv.cl

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_58

proposed situation the loss of customers is between 0 and 0.5%. In relation to costs and profits, these were obtained with model P, although the costs increase by about 5% since a larger inventory is maintained than in the current situation, but customers will be 100% satisfied, which leads to an increase in profits of between 15 and 20%.

Keywords Inventory management · Simulation · BPM · Neural networks

1 Introduction

Companies need to procure goods and services for the development of their activities. These supplies accumulate in the companies and must be managed for their correct handling and conservation. Small companies must have sufficient competencies and skills to permanently improve the management of their resources and thus remain inserted in the market.

For companies to have a positive impact on customer satisfaction, it is important to have an adequate inventory management to meet the demand that arises, avoiding stock outs that hinder the sales process. In addition to providing excellent customer service, differentiating the company from the competition from the moment the customer arrives at the store to after-sales service.

Currently, due to the effect of the pandemic, the company under study, began to work online, e-commerce, so every time inventory management is necessary for an improvement in customer service, since in several cases, customers, when there is no product in store these are lost, sometimes customers must wait and return the day the product is in the store.

For the above reasons, the objective of this work is to propose improvements to inventory management with the Business Process Management (BPM) [1] simulation tool, Bizagi Modeler for a small jewelry company (SME).

Mathematical models were also applied to define inventory management policies in order to determine quantities to order, how often to order, along with offering improvement options to provide a better level of service to the customer and evaluate the economic proposal, comparing it with the costs of the current situation. At the same time, the mathematical model of periodic inventory review will be compared with the simulation in Bizagi, analyzing quantities to order and associated costs.

The jewelry store specializes in marketing luxury items such as 18 k and 24 k gold jewelry, mainly rings, both in physical and online stores. It is noteworthy that according to the Chilean Ministry of Economy SMEs in Chile are 48.7% of companies [2].

2 Bibliographic Review

According to [3–6], the added value that logistics incorporates in each part of the logistics system is an important competitive weapon. Such is the case of excellence in delivery service, leadership in product differentiation, management at minimum cost or logistics service to the customer based on efficient inventory management.

According to [7] large companies recognize well the benefits of SCM, but small and medium enterprises (SMEs) lag behind in appreciating how the integrated supply chain drives remarkable changes in business processes and work with positive results in better quality services, cost reduction and efficiency. Specifically, SMEs in Zimbahue as well as in Chile do not have sufficient knowledge about SCM and underestimate the potential benefits of SCM, given the above, this paper studies SCM for inventory management.

As well as [8] where says that large companies are well recognized the benefits of SCM, but small and medium enterprises (SMEs) are lagging behind in appreciating how integrated supply chain drives remarkable changes in business processes and work with positive results in better quality services, cost reduction and efficiency.

According to [9], SMEs identify contextual areas that can guide them to successfully use e-commerce and realize its potential benefits. On the other hand, e-commerce is said to have a positive impact on performance and the impact at the strategic level is indirect, mediated through operational and management levels.

In [10] inventory management must be given the attention it deserves for the company to remain competitive, flexible to demand and at low cost. Inventories are difficult to manage and control, and inventory managers find it difficult to know when to order and how much to order.

In [11] also used, the BPM methodology that allowed him to simulate the times and cost of each of the inventory management processes, of an SME company in Guayaquil, Ecuador, which allowed him to document all the processes of the distribution center of the company achieving the optimization of its resources.

On the other hand [12] says that flowcharts have been used worldwide as a graphical tool to describe the sequence of activities that correspond to the processes, in addition they bring benefits such as minimization of errors, related to the protocols of execution of activities in companies. They also model processes to improve inventory management with Bizagi Modeler 8.

3 Methodology

First, the company's historical sales data for the years 2017, 2018 and 2019 are compiled and classified using the ABC methodology to sort the items according to those that generate the most revenue, so as to group and work with the information.

Once the products that generate the most income are obtained, which are 5 types of 18 k carat rings, the SCM of the SME is modeled, showing the process of sales,

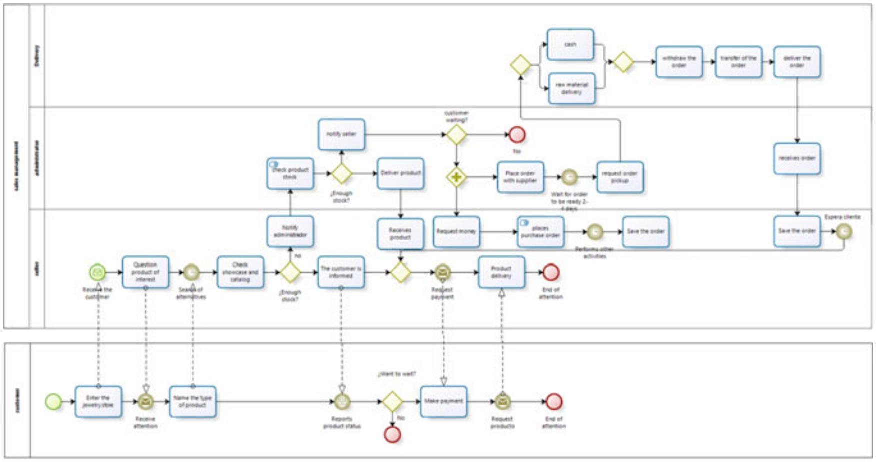


Fig. 1 Sales management process

inventory and request of raw materials, designed with Bizagi Modeler where the breakdowns are obtained, to improve them.

The current process design is shown in Fig. 1.

Figure 1 shows that the process has three participants: salesperson, replenisher and administrator. Each of these actors has specific activities to perform. The process begins when the salesperson serves a customer and ends when the customer purchases an available product or withdraws due to lack of stock.

The process has several breaks, where the main one is in the circle of Fig. 1, which means that the store manager is who is involved in the break point activity associated with the review of stock to subsequently deliver the product or products requested to the seller, if they are available, otherwise the seller must be notified of the depletion of the product to send the article to the supplier depending on the customer's response, in this case it is possible that the customer wait or get lost.

It is simulated with Bizagi Modeler, to obtain the results of the current situation according to the model in Fig. 1, and then the proposed situation without breaks is simulated.

On the other hand, demand was estimated using neural networks. The input data are the monthly demand of each selected item for the years 2017, 2018 and 2019. And the output data are the forecasts, until January 2020.

Subsequently, an inventory management model, periodic P, was analyzed. In order to choose this model, first of all, an analysis of the demand for the products must be made, where all the products under study that the company markets have independent demand, since no product depends on the other in terms of sales or product manufacturing. Therefore, we will advance along this line to find adequate inventory models.

To determine whether the demand for the products is deterministic or probabilistic, the historical data will be subjected to a rough estimation based on the calculation of the mean and standard deviation of consumption during a specific period (e.g. monthly).

Once the estimation was done, the result was that the behavior of all products is probabilistic non-stationary demand.

4 Results

To carry out the simulation process, it is assumed that the entry of customers is distributed according to a Normal distribution, with an average of 12 min, that is, every 12 min a customer enters, with a standard deviation of 2, the process is simulated for a month, with a schedule of 8 h, 7 days a week.

The process of the current situation, with historical data, starts by requesting stock for a month, where the lead time is 1 week and orders are requested for 14 products in the month.

The simulation was carried out only for the products with the highest margin, the five types of rings, 18 carat gold products.

4.1 Simulation Results

To simulate the current process, about 10 simulations were performed, until the best solution was reached:

- For all types of rings, when 14 orders are placed with the supplier per month, 70% of the products ordered are fulfilled, while 30% are not fulfilled, resulting in the loss of customers.

  For the proposed situation, where several simulations are carried out to obtain the minimum number of products requested and not served, the following results are obtained;
- For all types of rings, 20 units are requested from the supplier per month, 100% of the products requested are fulfilled, while 0% are not fulfilled and therefore no customers are lost.

  The stock is increased per month, thus reducing the number of customers lost to practically zero.

4.2 Inventory Management Results with Model P

Using the coefficient of variation, which yields a value of 20.05%, which is greater than 20%, together with an analysis of the demand in the different periods of a year,

Table 1 The results of the model P

Product	Input data				Output data		
Product	Monthly forecasted demand	Lead time (L)	Standard deviation of demand (σd)	Service level (CSL) (%)	Safety inventory (SS)	Target inventory	Quantity to order
Ring 1	16	0.033	5.6	95	11	26	16
Ring 2	30	1	8.3	95	15	73	30
Ring 3	15	0.1	5.6	95	12	26	15
Ring 4	17	0.13	4.8	95	10	27	17
Ring 5	13	0.07	3.5	95	7	19	13

and also visualizing one year with respect to the other, it is determined that the demand for the representative ring is probabilistic like the rest of the products, that is, it has a random behavior and cannot be predicted exactly. For this reason, the results of the inventory models must be updated for each period in order to provide assertive results in accordance with the reality of each period.

4.3 Cost Analysis

It is observed that the cost increase in the simulation with Bizager Modeler is 10.2% higher compared to the current situation, which is offset by the additional 30% of sales that can be obtained with this cost increase.

Comparing model P with the current situation, a decrease in costs of 12.6% is proposed, covering the entire current demand.

It can be seen that the situation that has a lower amount of costs is with model P, this is because there are fewer orders within a period, in turn the cost of maintaining inventories is low, so that total costs decrease when using this inventory model.

5 Conclusions

In this paper we have presented a real inventory management problem in an SME. For which the Bizagi Modeler and the P management model are used, achieving the desired results. Both proposed models will help the company to maintain a monthly stock so as not to lose customers, thus producing an increase in sales, and thus decreasing customer losses.

As a conclusion in the situation proposed with the Bizage Modeler, by increasing the stock of the month, the number of lost customers is practically reduced to 0.

With the inventory management model P, it is determined that the demand of the rings is probabilistic, i.e. it has a random behavior and cannot be predicted exactly. For this reason, the results of this inventory model must be updated for each period in order to deliver assertive results in accordance with the reality of each period.

Regarding costs, in the simulation with Bizager Modeler they are 10.2% higher compared to the current situation, which is offset by the additional 30% of sales that can be obtained with this increase in costs. Comparing model P with the current situation, a decrease in costs of 12.6% is proposed, covering the entire current demand.

Company owners decide on the inventory model to apply depending on their budgetary policies.

In the future, it is expected to simulate other processes in the company, as well as to design a system that can be updated periodically.

References

1. Bizagi Modeler: User's guide (454–514) retrieved from: http://help.bizagi.com/process-modeler/en/ (2020)
2. Report of results: Companies in Chile Fourth Longitudinal Survey of Companies Studies Unit, Ministry of Economy Chile (2017). https://www.economia.gob.cl/wpcontent/uploads/2017/03/Bolet%C3%ADn-empresas-en-Chile-ELE4.pdf. 25 August 2020
3. Christopher, M.: Logistics and Procurement: How to Reduce Costs, Stocks and Improve Services. Editorial Folio, Madrid (1994)
4. Prida, B., Gutiérrez, G.: Procurement Logistics. McGraw-Hill, Madrid (1997)
5. Levy, D.: Lean Production in an International Supply Chain (1997)
6. Gómez, M., Acevedo, J.A.: Procurement Logistics. Corporación John F. Kennedy, Bogotá
7. Muchaendepi, W., Mbohwa, C., Kanyepe, J.: Inventory management and performance of SMEs in the manufacturing sector of Harare. Proc. Manuf. **33**, 454–461
8. Chin, T.A., Hamid, A.B.A., Rasli, A., Baharun, R.: Adoption of Supply Chain Management in SMEs. Int. J. Pro. Econ. **208**, 421–433 (2019)
9. Panca, O.H.P., Harry B.S.: Contextual factors and performance impact of e-business use in Indonesian small and medium enterprises (SMEs). Heliyon **6**(3), e03568 (2020)
10. Nemtajela N., Mbohwa C.: Relationship between Inventory Management and Uncertain Demand for Fast Moving Consumer Goods, Organisations. Procedia Manuf. **8**,699–706 (2017)
11. Silva: Modeling and documentation of the inventory management processes of the distribution center sunplastic cia. ltda. Thesis for the bachelor's degree in industrial engineering., Universidad de Guayaquil, Ecuador (2018)
12. Morocho: Implementation of flowcharts in activities performed in the mishell pharmacy using the bizagi symbology to improve execution times, thesis for the master's degree of the Universidad Tecnica de Machala, Ecuador (2019)

Investigation of Transport Logistics Disruptions from Urban Floods: A Case Study of the Chinese Coastal Megacity—Guangzhou, China

Xiaohui Lu, Faith Ka Shun Chan, Hing Kai Chan, and Wei-Qiang Chen

Abstract Chinese coastal megacities face an increasing challenge on urban floods in light of climate change and rapid urbanisation. Severe urban floods worsen damages of public infrastructures, such as roads and railway networks, because the land drainage system is currently insufficient to cope with intensive rainstorms. This condition is particularly severe in most Chinese megacities. As a result, that disrupted the transport system and logistics, which associated the interruption of economic and business opportunities. This paper took Guangzhou, the leading commercial and manufacturing hub in the Greater Bay Area region, as a case study. Unfortunately, the city is currently exposed to a high level of urban flood risk. By analysing the historical data, this research summarised the annual rainfall pattern in Guangzhou, and presented the potential impacts of the urban flood on logistics disruption based on the layout of road surface water flooding. Our findings in this research aim to provide recommendations on logistics disruption and transportation plan for road users in future flood events. Moreover, the results are insightful in tackling future urban floods to improve current policies on achieving resilient city planning on logistics.

Keywords Logistics disruption · Road networks · Climate change · Urban flood · Coastal megacity

X. Lu · H. K. Chan (✉)
Nottingham University Business School China, University of Nottingham Ningbo China, Ningbo 315100, China
e-mail: Hingkai.Chan@nottingham.edu.cn

F. K. S. Chan
School of Geographical Sciences, Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo 315100, China

Water@Leeds Research Institute and School of Geography, University of Leeds, Leeds LS2 9JT, UK

X. Lu · W.-Q. Chen
Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_59

1 Introduction

Infrastructure systems such as railway and road networks are the backbones and skeleton of cities, especially in the context of the development of transportation network [1]. In particular, road networks play a crucial role in the whole infrastructure system, underpinning the mobility of goods, logistics and people under both normal and emergency conditions [2]. As cities are the agglomerations of people and companies, road networks are extended rapidly for various human and manufacturing activities. Deficient performance of road networks can lead to severe disruption of the urban system [3]. For example, transport disruption causes the absence of daily commutes (e.g. going to work and school), interruption of economic and business opportunities, high cost of public services [2]. The performance would become worse under adverse and extreme flood events [4]. Flood hazard is the most frequent natural hazard in cities, which occurs as an overflow of water from water bodies (e.g. rivers, lakes and oceans), or occurs due to an accumulation of rainwater on the saturated ground [5].

According to the United Nations report, there are 3254 flood disaster events from 2000 to 2019, followed by 2043 storm disaster events and 552 earthquake disaster events [6]. Among China's top ten disasters in 2020, 4 disaster events are related to extreme rainfall and induced flood, resulting in 225.56 billion RMB direct economic losses [7]. There are three types of flood based on occurrence conditions: coastal, riverine, and urban floods. In this study, we focus on urban floods.

Urban flood (similar term as 'surface water flooding') has become a threat to the transport system in urban areas, where the permeability of the land surface and sufficiency of the drainage system has reduced. According to the existing studies, approximately 39% of the global land area suffers from urban flood [8]. Indeed, urban flood events have occurred frequently and intensely. In particular urban flood represents the most common flood hazard in the urban area [9]. The apparent phenomenon of an urban flood is the presence of surface water runoff on the impervious roads (e.g. highway, motorway) [10]. Changes in the natural (e.g. climate change) and anthropogenic (e.g. rapid urbanisation and developments) factors are escalating urban road networks under increasing pressure [11]. The transport disruption and reduction in performance of the transport system via floods could cost more than 900 thousand RMB per hour for each main road affected [12].

Urban flood impacts on transport disruption could directly affect the performance of the logistics industry. It is widely accepted that logistics companies' location has a strong dependence on the performance of transport infrastructure. Numerous logistics companies tend to locate around the trunk road, airport, port, railway station, etc. [13]. The inundation of these locations and roads would increase the goods delivery time and even lead to the goods damage. However, existing studies mainly focused on the impact of transport disruption on daily commutes caused by urban floods. With the increase of the extreme rainfall, it is a considerable demand to investigate the logistics disruption caused by urban flood in cities with significant manufacturing activities and high level of urban flood risk (e.g. Guangzhou, Shenzhen). This study is devoted to the investigation of urban-flood-driven impacts on logistics disruption. The detailed objectives are as follow: (1) analyse the annual rainfall in Guangzhou at both city level and district level; (2) report on changes in the layout of road waterlogging in 2020 and 2021; and (3) identify the potential impacts of urban floods on logistics disruption. Based on the findings, this paper recommends both academic and managerial insights for future studies.

2 Methodology

2.1 Case Study of Guangzhou

Guangzhou is a coastal megacity with 12.64 million people in Southern China. The city is the capital of the province of Guangdong, the leading manufacturing hub of the Pearl River Delta. Moreover, it is one of mainland China's leading commercial and manufacturing regions [14]. The logistics industry plays a crucial role in the economic development of Guangzhou. According to the national statistics, Guangzhou's express service volume and income ranked second in China in 2020 [15]. Guangzhou is prone to urban floods due to the large proportion of impervious surface (e.g. roads, buildings, etc.) and inadequate drainage systems. Land use has rapidly changed since the economic reform and open-door policy in 1978 [16, 17]. In the past five years, numerous rainstorms hit Guangzhou and led to severe damages. For example, urban flood events led to 8 deaths, and one metro line flooded in Guangzhou in 2016. The rainstorm (wide range intense rainfall) led to about 135 waterlogging spots in the central city, of which 26 had a waterlogging depth of more than 40 cm on 19–20 April 2019. The torrential rain caused widespread transportation disruption (see Fig. 1) and claimed 4 lives on 22 May 2020, with a direct economic loss of 800 million yuan [18].

Fig. 1 Road waterlogging in Guangzhou during the rainfall season. *Source* [18]

2.2 Data and Analysis

Statistical analysis is used to analyse the rainfall characteristics in Guangzhou both at a city level and district level. The rainfall data used in this study was extracted from Guangzhou Climate Bulletin 2016–2020 published by the Meteorological Bureau of Guangzhou Municipality (http://www.tqyb.com.cn). Guangzhou Climate Bulletin 2016–2020 provides major urban flood events and damages in the last five years. For example, a rainstorm hit Guangzhou on 19 April 2019, resulting in 135 roads waterlogging [19]. There are various standards to define rainstorm conditions and urban flood hazards. Guangzhou Climate Bulletin classifies rainstorm into three types: rainstorm (rainfall > 50 mm in 24 h), heavy rainstorm (rainfall > 100 mm in 24 h), and torrential rainstorm (rainfall > 250 mm in 24 h). There are five classifications on urban floods: no urban flood (rainfall < 25 mm in 24 h), light urban flood (rainfall > 25 mm in 24 h), common urban flood (rainfall > 50 mm in 24 h), severe urban flood (rainfall > 100 in 24 h), and catastrophic urban flood (rainfall > 250 mm in 24 h), which are adopted to determine the potential urban flood hazards [14]. A list of 69 logistics parks was obtained from China Wutong Website (https://www.chinawutong.com/baike/). The information of waterlogging spots promulgated by Guangzhou Water Authority (http://swj.gz.gov.cn) was combined with the geographic information of logistics parks to investigate the potential urban flood impacts on logistics disruption.

3 Result and Discussion

3.1 The Annual Rainfall from 2016 to 2020

The average rainfall of Guangzhou was 2150.9 mm during the last five years. As shown in Fig. 2, the 2nd quarter was the most abundant rainfall quarter, followed by the 3rd quarter. Rainfall of the two quarters accounted for 35%–55% and 29%–40% of annual rainfall, respectively. It can be seen that heavy rainstorm in the 2nd and 3rd quarters and severe drought in the 1st and 4th quarters are the major characteristics in Guangzhou. The causes of heavy rainfall in these quarters are the dragon boat water and typhoon, respectively. Figure 2 shows that severe drought in the 4th quarter is becoming more evident over the years. For example, the rainfall of the 4th quarter is 222 mm in 2016, while it is only 20 mm in 2020. From the perspective of spatial distribution, the wettest district is Huangpu district (2253 mm), followed by Zengcheng district (2189 mm) and Baiyun district (2120 mm). The monitoring and warning systems should be improved to protect the three districts from urban flood due to they are more vulnerable and exposed to the rainstorm. The current warning systems play a significant role in protecting people and assets from urban flood prior to rainstorms. There are 67 yellow rainstorm warnings, 13 orange rainstorm warnings and 2 red rainstorm warnings from 2018.

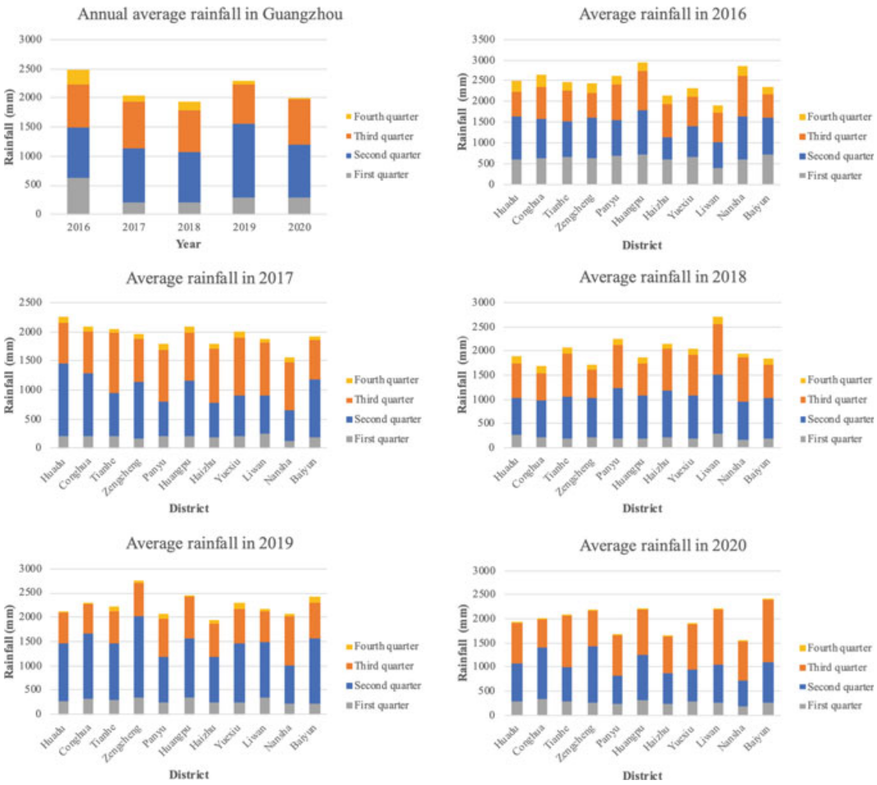


Fig. 2 Annual rainfall of different districts in Guangzhou from 2016 to 2020. *Source* [18–22]

3.2 *The Layout of Road Waterlogging*

Based on the analysis at a city level, it can be seen that the number of waterlogging spots has decreased from 128 in 2020 to 93 in 2021. As shown in Fig. 3, the number of waterlogging spots in the Zengcheng district has significantly reduced from 30 in 2020 to 11 in 2021. The maximum number of waterlogging spots in 2021 is 13, located in the Baiyun district and the Tianhe district. Heavy rainfall and insufficient drainage system are the leading factors that led to waterlogging in these districts. For example, the depth of water in some roads was over 100 cm and even over 300 cm extracted from the historical flood events in Guangzhou previously.

We analysed the historical data extracted by the Guangzhou Meteorological Bureau, which has recorded the 100 cm water depth or above that occurred on the Huangyuan road in the Baiyun district during 13 September 2010. Unfortunately, a surface water flood occurred at 300 cm water depth on Xinshi road in the Zengcheng district on 7 May 2018 [20].

However, the current urban flood protection flood standard remains inadequately adapted according to the local Master Plan and district plan. The existing drainage

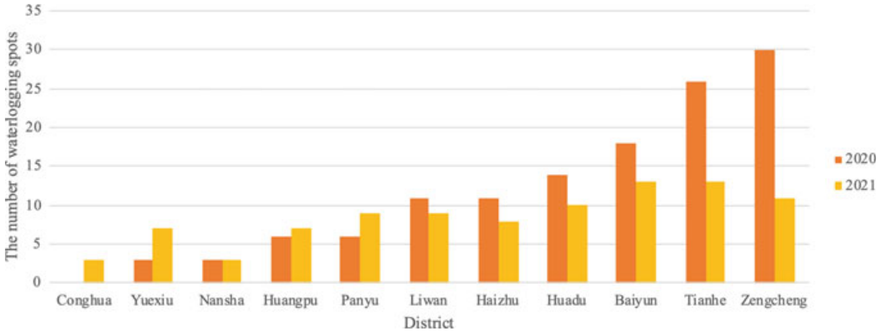


Fig. 3 The number of waterlogging spots in different districts in Guangzhou. *Source* [24]

protection standard in the metropolitan and municipality urban areas in Guangzhou remains inadequately low with the 1-in-1-year urban flood. Drainage systems in these areas must be reconstructed to resist the 1-in-2-year to 1-in-3-year urban flood. For new roads, the drainage protection standard is 1-in-5-year to 1-in-10-year return period [23].

3.3 The Potential Impacts of Urban Floods on Logistics Disruption

Table 1 shows the number of logistics parks in different districts. There are 33 parks in the Baiyun district due to the developed road networks and Guangzhou Baiyun International Airport, and most of them tend to locate around the highway. For example, seven logistics parks located around Dayuan North Road of the Baiyun district, which is the 1st-grade highway with 60 km/h–100 km/h vehicle speed [25]. The number of logistics parks in Huangpu is ranked second due to its advantageous shipping transport and road network.

Traffic condition is the primary factor for the distribution of logistics parks. However, traffic condition is adversely affected by heavy rainfall and increasing road inundation. It can be seen that transport disruption caused by an urban flood is a frequent and stubborn problem in areas with heavy rain and high traffic pressure. Based on the urban flood impacts on transport disruption [26], the potential impacts of urban floods on logistics disruption can be summarised into three types: logistics delay, driver's safety, and goods damage. Road inundation is the first phenomenon after heavy rain and the direct cause of traffic congestion. Logistics would be delay or event disrupted caused by the traffic congestion. Delivery schedules could be adversely affected by the increasing depth of road inundation. Moreover, heavy rain and induced urban flood often led to serve damages to goods. For the general goods, dry conditions are needed for storage and transportation. For fresh goods, timeliness and cold conditions are significant [27].

Table 1 The number of logistics parks in different districts in Guangzhou

District	Conghua	Yuexiu	Nansha	Huangpu	Panyu	Liwan	Haizhu	Huadu	Baiyun	Tianhe	Zengcheng
Number	1	0	0	16	3	2	3	5	33	0	6

Source [28]

Combining the analysis of the waterlogging layout, logistics parks in the Baiyun district and the Zengcheng district need to be paid more attention to the transport logistics planning and rerouting to avoid numerous waterlogging spots during the heavy rain conditions. From the perspective of inundation time and depth, 14 waterlogging spots were flooded for more than 1 h, even reached 9 h in the Huangpu district, such as the Kaiyuan Avenue Tunnel, Shihua road-Wencheng road. These roads in the Huangpu districts need to be considered for logistics location and route selection. Moreover, strengthening urban flood management and logistics planning is essential to enhance logistics performance during urban flood conditions.

4 Recommendations

4.1 Quantify Logistics Disruption Using a Simulation Model

Quantifying the impacts of urban floods on logistics disruption can provide a better understanding to enhance logistics performance. For example, the adequate time of traffic delay on the given road sections is valuable information for decision making (i.e. rerouting) under urban flood conditions. There is a considerable demand for quantifying the logistics disruption in the areas with increasing manufacture and logistics activities [29]. Rainfall and traffic are two leading factors that contribute to the logistics disruption caused by urban floods. Simulation approach was widely adopted to evaluate the relationship between rainfall and traffic delay. Flood model and traffic model can be used to simulate the potential road inundation and traffic delay.

For example, Li et al. developed an integrated methodology for modelling the traffic disruption caused by urban flood in Shanghai [30]. Geographical Information Systems (GIS) can be integrated to analyse the spatial distribution of logistics companies and map the logistics disruption. The findings of simulation and spatial analysis could help determine the critical locations and prioritise the investment and services before or during urban flood events. Therefore, the combined use of flood model, traffic model and GIS analysis is recommended to quantify the impacts of urban floods on logistics disruption at a city level.

4.2 Practices and Planning

From 'grey' to 'green' and 'blue' strategies, several adaptation strategies have been proposed for reducing urban area prone to urban flood. The sponge city initiative is essential for China's cities to enhance water sensitivity and urban flood resilience

[31]. Although Guangzhou is not a pilot city for the sponge city program, city authorities have produced the Guangzhou Sponge City special planning (2016–2030) initiative in 2017 [23]. The planning suggested that 51 sponge parks need to be renovated or built to improve the capacity of urban flood resilience.

From the perspective of indirect impacts, the cooperation between the institutions (e.g. municipality bureaus and departments—i.e. transport, water, etc.) is important to alleviate the according impacts of transport disruption that caused by urban floods. For example, integrating the meteorological information and traffic information could provide a better understanding for protecting people and assets from the potential urban flood. Guangzhou Municipality authorities should open some channels and encourage citizens to report the new waterlogging spots.

With the increase of annual rainfall, logistics companies should pay more attention to transport disruption. For example, they can develop the flood module to display the vulnerable roads and alternative roads in their information platform or system. Adaptation of urban flood is recommended to be involved in logistics company's strategies to enhance their logistics performance. In addition, an in-deep consideration of insurance is significant for these logistics companies located in heavy-rainfall areas. Therefore, the practices of urban flood management need the contribution from city authorities, companies and citizens.

5 Conclusion

This paper presents an investigation of urban flood impacts on logistics disruption. Guangzhou, the leading commercial and manufacturing hub in the Greater Bay Area region, was taken as the case study due to its high level of urban flood risk. Regarding the whole city of Guangzhou, the heavy rainstorm in the 2nd and 3rd quarters and severe drought conditions in the 1st and 4th quarters are the major characteristics in Guangzhou. Moreover, the phenomenon of severe drought in the 4th quarter is becoming more obvious over the years. Logistics companies need to pay attention to reduce the impacts of urban floods (i.e. logistics delay, driver safety and goods damage), especially in the highly vulnerable and exposed districts.

In this study, based on the patterns identified in Guangzhou, the combined use of flood model, traffic model and GIS analysis is recommended to quantify the impacts of urban floods on logistics disruption at both a city level and district level. From the perspective of practice and planning, the cooperation of city departments is recommended for urban flood management. In future, we would employ simulation models to evaluate the impacts of urban floods on logistics disruption. We suggested that the authorities should identify and provide more information on the according vulnerable areas of urban floods for the road users and that will be increasing the urban flood resilience and improving the transport logistics efficiency under adverse flood conditions in Guangzhou and other Chinese cities.

References

1. Rodrigue, J.-P., Notteboom, T.: Transportation and economic development. J. Geogr. Transp. 1–29 (2013)
2. Rozenberg, J., Marianne, F.: Beyond the gap: how countries can afford the infrastructure they need while protecting the planet. The World Bank (2019)
3. Hernandez-Fajardo, I., Dueñas-Osorio, L.: Probabilistic study of cascading failures in complex interdependent lifeline systems. Reliab. Eng. Syst. Saf. **111**, 260–272 (2013). https://doi.org/10.1016/j.ress.2012.10.012
4. Koetse, M.J., Rietveld, P.: The impact of climate change and weather on transport: an overview of empirical findings. Transp. Res. Part D Transp. Environ. **14**(3), 205–221 (2009). https://doi.org/10.1016/j.trd.2008.12.004
5. Plate, E.J.: Flood risk and flood management. J. Hydrol. **267**(1–2), 2–11 (2002). https://doi.org/10.1016/S0022-1694(02)00135-X
6. UN Office for Disaster Risk Reduction: Human cost of disaster: an overview of the last 20 years: 2000–2019. https://reliefweb.int/report/world/human-cost-disasters-overview-last-20-years-2000-2019 (2020)
7. Ministry of Emergency Management of the People's Republic of China: China's Top Ten Disasters in 2020. https://www.mem.gov.cn/xw/yjglbgzdt/202101/t20210102_376288.shtml (2021)
8. Koks, E.E., Rozenberg, J., Zorn, C., Tariverdi, M., Vousdoukas, M., Fraser, S.A., Hall, J.W., Hallegatte, S.: A global multi-hazard risk analysis of road and railway infrastructure assets. Nat. Commun. **10**(1), 2677 (2019). https://doi.org/10.1038/s41467-019-10442-3
9. Kendon, E.J., Roberts, N.M., Senior, C.A., Roberts, M.J.: Realism of rainfall in a very high-resolution regional climate model. J. Clim. **25**(17), 5791–5806 (2012). https://doi.org/10.1175/JCLI-D-11-00562.1
10. Adedeji, T., Proverbs, D., Xiao, H., Cobbing, P., Oladokun, V.: Making Birmingham a flood resilient city: challenges and opportunities **17** (2019)
11. Jiang, Y.: Urban pluvial flooding and Stormwater management—a contemporary review of China's challenges and 'sponge cities' strategy **12** (2018)
12. Pregnolato, M., Ford, A., Glenis, V., Wilkinson, S., Dawson, R.: Impact of climate change on disruption to urban transport networks from pluvial flooding. J. Infrastruct. Syst. **23**(4), 04017015 (2017). https://doi.org/10.1061/(ASCE)IS.1943-555X.0000372
13. Rao, C., Goh, M., Zhao, Y., Zheng, J.: Location selection of city logistics centers under sustainability. Transp. Res. Part D Transp. Environ. **36**, 29–44 (2015). https://doi.org/10.1016/j.trd.2015.02.008
14. Lyu, H.-M., Wang, G.-F., Shen, J., Lu, L.-H., Wang, G.-Q.: Analysis and GIS mapping of flooding hazards on 10 May 2016, Guangzhou, China. Water **8**(10), 447 (2016). https://doi.org/10.3390/w8100447
15. State Post Bureau of the People's Republic of China: Statistical bulletin on the development of postal industry in 2020. http://www.spb.gov.cn/xw/dtxx_15079/202105/t20210512_3901027.html (2021)
16. Chen, M., Huang, Y., Tang, Z., Lu, D., Liu, H., Ma, L.: The provincial pattern of the relationship between urbanisation and economic development in China. J. Geog. Sci. **24**(1), 33–45 (2014)
17. Huang, L., Yan, L., Wu, J.: Assessing urban sustainability of Chinese megacities: 35 years after the economic reform and open-door policy. Landsc. Urban Plan. **145**, 57–70 (2016)
18. Meteorological Bureau of Guangzhou Municipality: Guangzhou Climate Bulletin 2020. http://www.tqyb.com.cn/gz/climaticprediction/bulletin/2021-02-07/7956.html (2021)
19. Meteorological Bureau of Guangzhou Municipality: Guangzhou Climate Bulletin 2019. http://www.tqyb.com.cn/gz/climaticprediction/bulletin/2020-03-27/6231.html (2020)
20. Meteorological Bureau of Guangzhou Municipality: Guangzhou Climate Bulletin 2018. http://www.tqyb.com.cn/gz/climaticprediction/bulletin/2019-01-18/4218.html (2019)
21. Meteorological Bureau of Guangzhou Municipality: Guangzhou Climate Bulletin 2017. http://www.tqyb.com.cn/gz/climaticprediction/bulletin/2018-02-12/2845.html (2018)

22. Meteorological Bureau of Guangzhou Municipality: Guangzhou Climate Bulletin 2016. http://www.tqyb.com.cn/gz/climaticprediction/bulletin/2017-01-23/1370.html (2017)
23. Guangzhou Municipal Planning and Natural Resources Bureau: Guangzhou Sponge City Special Planning (2016–2030) (2017)
24. Guangzhou Water Authority. Homepage. http://swj.gz.gov.cn
25. The Ministry of Housing and Urban-Rural Development of the People's Republic of China: Code for Design of Urban Road Engineering. CJJ 37-2012 (2016)
26. Pregnolato, M., Ford, A., Glenis, V., Wilkinson, S., Dawson, R.: Impact of climate change on disruption to urban transport networks from pluvial flooding. J. Infrastruct. Syst. **23**(4), 04017015 (2017)
27. Liu, H., Pretorius, L., Jiang, D.: Optimisation of cold chain logistics distribution network terminal. EURASIP J. Wirel. Commun. Networking **2018**(1), 158 (2018). https://doi.org/10.1186/s13638-018-1168-4
28. China Wutong Website. Homepage. https://www.chinawutong.com/baike/
29. Garrido, R.A., Lamas, P., Pino, F.J.: A stochastic programming approach for floods emergency logistics. Transp. Res. Part E Logistics Transp. Rev. **75**, 18–31 (2015). https://doi.org/10.1016/j.tre.2014.12.002
30. Li, M., Huang, Q., Wang, L., Yin, J., Wang, J.: Modeling the traffic disruption caused by pluvial flash flood on intra-urban road network. Trans. GIS **22**(1), 311–22 (2018). https://doi.org/10.1111/tgis.12311
31. Chan, F.K.S., Griffiths, J.A., Higgitt, D., Xu, S., Zhu, F., Tang, Y.-T., Xu, Y., Thorne, C.R.: Sponge City in China—a breakthrough of planning and flood risk management in the urban context. Land Use Policy **76**, 772–778 (2018). https://doi.org/10.1016/j.landusepol.2018.03.005

Locating Street Markets in Smart Cities: A View from Mathematical Modeling and Urban Planning Perspectives

Gonzalo Mejía, Daniela Granados-Rivera, and Carolina Avella

Abstract Locating facilities has traditionally viewed from different disciplines. However, few studies are interdisciplinary. In this paper, we investigate the location of street markets in smart cities from two perspectives: urban planning and mathematical modelling. Smart city technologies can play a major role in the implementation of street markets: for example, smart street markets can show in real time the availability of products, market prices, etc. In this paper, we propose a mathematical model to locate "smart" street markets in the city of Bogotá (Colombia) with constraints that involve urban planning and public space. The results will show how this strategy can improve the current practice.

Keywords Street markets · Facility location · Urban planning

1 Introduction

Street markets have existed for many centuries around the world. They provide not only fresh food and other products but also, they are part of the social network in many places around the world. They have survived waves of competitors and still today they are very important in both the developed and emerging countries. Street markets normally are hosted by local sellers whose stalls that set-up and disassembled in the same day. In general, street markets provide a large assortment of products, some normally not found in other retailers such as corner stores and supermarkets. In addition, as these street markets are attended by local producers and do not rely

G. Mejía (✉) · D. Granados-Rivera
Faculty of Engineering, Universidad de La Sabana, Campus Universitario Puente del Común Km7. Autopista Norte de Bogotá D.C., Bogotá, Colombia
e-mail: gonzalo.mejia@unisabana.edu.co

D. Granados-Rivera
e-mail: danielagrri@unisabana.edu.co

C. Avella
School of Architecture, Fundación Universitaria de Colombia, Bogotá, Colombia

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_60

on intermediation, they can offer lower prices in comparison with their competitors. Needless to say that, street markets also have disadvantages such as pollution, land degradation, and problems with waste disposal [1].

The location of street markets has been traditionally established by public authorities based on empirical and political considerations [2]. Usually, these considerations do not consider many important aspects such as road access, parking areas, land use, population density and urban design. Locating street markets is an interdisciplinary challenge that has a major impact on their success or failure as pointed out by [2]. From the engineering and business perspective, locating street markets aims primarily at establishing the locations where the market share is maximized. Normally this is done with discrete choice models [3] where the total "attractiveness" is maximized. Attractiveness is a measure of the utility that customers perceive of a facility based on attributes such as price, distance, assortment, and aesthetical considerations. However, since street markets are public institutions, ingrained in the society and part of the urban landscape, aspects related to urban planning must not be neglected. In fact, a holistic view is indeed needed for an appropriate planning.

A major difficulty is to combine such quantitative engineering methods with qualitative approaches of social and urban planning disciplines. In this paper, we aim to provide a first approximation for such this holistic view with some insights about their implementation in smart cities.

2 Background

2.1 Competitive Facility Location

The location of street markets (also known as farmers' markets, open-air markets, mobile markets) is an uncommon topic in the literature. Although many authors have demonstrated their benefits on food insecurity problems (e.g. [4, 5], there are only three studies that employed mathematical models to define the optimal locations done by [6–8]. However, none of these authors have used a competitive approach that is an important point by considering customer choice and some market behavior features. The competitive facility location is a variation of the traditional facility location problem that implements a choice constraint based on discrete choice models [9–11] or gravity models [12, 13]. Usually, the most common factors are price, distance, and travel time.

However, these factors rarely consider urban planning features to evaluate the attractiveness of a location according to its infrastructure.

2.2 Selection of Suitable Locations: Urban and Functionality Factors

Locating a street market in a public area implies taking into consideration not only the economic feasibility of mathematical and location models, but also the urban factors that influence whether the spaces chosen for these markets are appropriate. The urban planning features used to evaluate the locations were chosen according to the criteria for locating facilities and farmers markets according to [14, 15]. These include demographic (population, social stratification), the functionality of the market (vehicle's road width, sidewalks and hard floor, morphology) and the area attributes (social facilities, land use, public transport and traffic).

In this paper, we use the classical weighted linear combination of factors to preselect a subset of sites based on the recommendations of [14, 15] and then we use a mathematical model to maximize the market share. Details follow.

2.3 Smart Cities

The concept of smart city is rooted in the need to find new ways to improve the human life quality in a sustainability manner and with massive use of information technology [16, 17]. Smart cities can help to improve the efficiency of street markets by providing ways to display prices and product availability in real time.

3 Methods

This section introduces the methods used in this study in two parts: (i) the mathematical model employed to locate street markets, and (ii) criteria to preselect the potential location of street markets according to different urban planning features.

3.1 Competitive Facility Location Model

We employed a basic model of the competitive facility location to determine the street market sites. Our model considers a set of household clusters $j \in J$ which can choose to fulfill their average daily demand h_j among four types of sellers $s \in S = \{I \cup K \cup F \cup M\}$: (i) street markets $i \in I$, *corner stores* K, *fruit and vegetable sellers (fruver)* F, and supermarkets M. The probability y_{sj} that a household selects any of these sellers $s \in S$ is given by a discrete choice model, as follows:

$$y_{sj} = \frac{e^{-c_{sj}}}{\sum_{s' \in S} e^{-c_{s'j}}} \tag{1}$$

in which c_{sj} is the perceived cost function of selecting the seller $s \in S$ by household $j \in J$. This cost function contemplates, as main factors, the distance between the household and seller location d_{sj} and the selling price of that seller p_s. Each of these factors has a weight β_1 and β_2 respectively according to its impact on the buying decision. According to some authors (e.g. [1]), we selected both factors since price and distance are the first and second most important factors in the market choice.

Under this regard, we aim to select street market locations $i \in I$, which maximize the demand captured measured by the buying probability in street markets and the average daily demand of households. This location problem modifies Eq. (1) according to the open street markets. We assume that each street market has a daily capacity Q_i and that we have a limit of street markets Max that can be open. With this, the mathematical formulation is shown below.

Decision variables

$$X_i = \begin{cases} 1 & \text{If street market at i is located} \\ 0 & \text{Otherwise} \end{cases}$$

y_{ij} = Probability of selecting street market i by household j

As we mentioned before, the objective function is to maximize the market share, which is the product of the purchasing probability and the average daily demand as presented in Eq. (2).

$$Max \sum_{i \in I} \sum_{j \in J} h_j y_{ij} \tag{2}$$

The constraints are showing below. Constraint set (3) refers to the discrete choice model to determine the buying probability according to the street markets open. Constraint set (4) defines the parameter of the cost used in the discrete choice model. Constraint set (5) avoids that a street market sells more than its capacity. Constraint set (6) limits the number of street markets to open. Finally, constraints set (7) and (8) correspond to the nonnegative and binary restrictions for variables decision.

$$y_{ij} = \frac{X_i e^{-c_{ij}}}{\sum_{s \in I} X_s e^{-c_{sj}} + \sum_{s \in S/I} e^{-c_{sj}}} \quad \forall i \in I;\ j \in J \tag{3}$$

$$c_{sj} = \beta_1 d_{sj} + \beta_2 p_s \quad \forall s \in S;\ \ j \in J \tag{4}$$

$$\sum_{j \in J} h_j y_{ij} \le Q_i X_i \quad \forall i \in I \tag{5}$$

$$\sum_{i \in I} X_i \leq Max \tag{6}$$

$$X_i \in \{0, 1\} \quad \forall i \in I \tag{7}$$

$$y_{ij} \in [0, 1] \quad \forall i \in I;\ j \in J \tag{8}$$

Considering that Eq. (3) is a nonlinear function, we used the linearization proposed by [18] for multinomial logit choice probabilities.

3.2 Preselection of Potential Location

As said above, we use the weighted linear combination of factors to establish the attractiveness of a potential location from the urban planning. The weights were established with an expert in urban planning. Those with the highest incidence are those related to land use and morphology, which were assigned a weight of 20 points. According to city regulations (POT),[1] there are uses compatible with housing such as local-scale businesses, which would include street markets. The morphology of the public space is another of the determining factors with greater incidence, since in parks with slopes, low vegetation and continuous enclosure (fences) are not appropriate for a street market. Population or customer base is the next highest score with a 15 points weight.

The next three features with a medium incidence (10 points) are the social stratification, the vehicle road width and the sidewalks or available hard floor. The social stratification is associated to the need of bringing fresh food to low-income neighborhoods, and when a street market is located in a public area next to 1, 2 or 3 strata,[2] it will be an option for this population to get fresh food. Road width is related to the farmers requirements, to have a parking area next to the stands to load the groceries and the tents [15]. The need of having a flat hard floor area or/and a sidewalk of certain width is for the tents and the pedestrian flow.

Finally, the last three features with low incidence (5 point) are de social facilities in the location area, the public transportation and the streets with high traffic, associated to a better pedestrian flow as well as customers as it is well known. In future, models aspects such as Internet connectivity and information technology will be considered. We visualize this as a network of street markets where customers can place orders online based on real time information, pay and pick up their purchase. Public authorities can also benefit from this initiative by capturing data for quick network reconfiguration and to connect farmers and customers.

Table 1 summarizes the weights. The public areas scoring over 40 points are the options to locate the street markets that feed the model.

[1] POT: Plan de Ordenamiento Territorial, Decreto 190 de 2004 para la ciudad de Bogotá.

[2] In Bogotá the low-income population is usually located in 1, 2 and 3 strata.

Table 1 Factor weight

Factor	Weight
1. Population	15
2. Social stratification	10
3. Vehicle's road width	10
4. Sidewalks and hard floor	10
5. Shape morphology	20
6. Social facilities	5
7. Land use	20
8. Public transportation	5
9. Traffic	5
Total (Maximum score)	**100**

4 Description of Case Study

The area selected for this study was Engativá in the city of Bogotá, Colombia. Engativá is the 10th administrative division of Bogotá, composed of 405 neighborhoods of the low and lower middle classes in the northwest of the town. The population in this area is 814,000 [19]. For this research, we aggregate all the people in 500 clusters with an average of 2300 inhabitants. We consider a fruit and vegetable demand of 175.2 g on average per person [20]. According to the public park's database, the 52 initial street market potential location were selected and filtered using the available area to locate a street market [21]. We named this configuration Scenario 1. Other location retailers such as *nanostores*, fruvers (a type of grocery in which only sells fruits and vegetables), and supermarkets were found using Google Maps™. Their prices were taken from [22, 23]. The distances between sellers and demand points (households) were established using the Haversine formula [24]. Finally, we calibrated the model according to the figures in [25, 26].

In the town of Engativá there are multiple public spaces for recreational use. Within these public spaces, the 52 parks with the largest area, were preselected as optional location for a street market. Then each one was inspected with satellite images (Google Earth) and with the cartographic information on land uses from the POT [14] of Bogotá. This inspection identified its morphological characteristics, land use characteristics, the mobility structure (surrounding streets), the surrounding facilities, the public transport routes (associated with the bus stops) and the vehicular traffic in the area. These factors were combined with information on the population and social stratification to evaluate them and limit public spaces to those that, according to their urban conditions, allow the operation of the Street market and the influx of customers. Due to lack of space, we do not include the results of the preselected sites.

For the preselection of the potential street market locations, we employed socioeconomic and demographic information from public databases [27, 28]. First, we defined a buffer of 300 m around each street market location to establish the urban

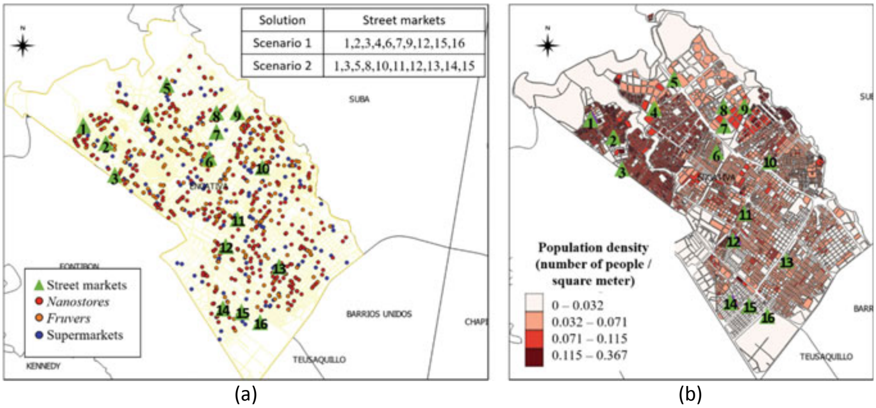


Fig. 1 Geographical distribution of the selected street markets for a solution with a maximum of 10 markets to open about **a** other sellers and **b** population density

planning features. Then, within those buffers, we described the strata classification of the point (median income level of households to determine rates of public utilities [29, 30]), the number of potential buyers (population in the buffer), and the number of bus stops as a measure of circulation and congestion of each point. Also, we used information about the infrastructure of each location to qualify that factor. With this preselection, the potential locations of street markets were reduced to 27 points shown in Fig. 1b. We called this configuration Scenario 2.

5 Results

In this section, we compare the results of the mathematical model of both scenarios: (1) without the preselection of the potential locations of street markets (scenario 1), and (2) with the preselection of the potential locations (scenario 2).

Initially, we vary the maximum number of street markets to locate for both scenarios. The results show that the market share grows in linear fashion. Evidently, the more street markets, the more percentage of demand captured. The increase by adding new locations tends to be linear, showing that the selected location's attractiveness is similar by capturing the same quantity of kilograms.

To compare the changes in the configuration of the street markets without and with the preselection according to the urban planning factors, we select the solutions with a maximum of 10 street markets to open in both scenarios. These geographical distributions are illustrated in Fig. 1a. We can observe that locations 1, 3, 12, and 15 are the only conservated from the solution without the preselection. It indicates that the remaining locations are not attractive to locate street markets from the features of urban planning contemplated in the preselection. These locations might theoretically capture the same percentage of the others in the solution with the preselection, but to real applications, those remaining locations do not have the infrastructure to be suitable for locating street markets.

From the competitive approach in both scenarios, we can analyze locations such as 7, 8, 9, and 16 in areas with a low number of retailers (see Fig. 1a), which shows that in this area, people will likely buy in street markets due to the lack of competition from other sellers. But locations such as 1, 2, 3, 6, 10, and 13 are in areas with many retailers making, initially, no sense why these markets are located there because there is more competence (see Fig. 1a). However, when we review the demand around each of these points, we can notice that these areas have a high population density (see Fig. 1b). It means that regardless of the number of sellers around, the demand capture is greater than in other areas explaining their selection.

6 Conclusions and Future Works

The objective of this study was to analyze the changes in the configuration of a network of street markets according to a preselection done using urban planning features. For this purpose, we used a competitive facility location model to maximize the market share. In the case study employed, we observed that the configuration does not affect the total percentage of the demand captured because the attractiveness of the markets, measured by the cost function, was similar. This is taken into account that the factors of the utility function are distance and price. However, by preselecting the initial locations, we realized that many of the selected locations in scenario 1 were unsuitable according to an urban planning approach. It is probably that these locations do not have the desirable infrastructure to locate a street market in terms of unloading fresh food of trucks, installing the stalls, soil features, people circulation, among others. For future research, implementing these types of characteristics in the cost function could ensure the real potential of a location evaluating urban planning constraints that are unusual considered.

References

1. Neckerman, K.M., Bader, M., Purciel, M., Yousefzadeh, P.: Measuring food access in urban areas. Citeseer (2009)
2. Rajagopal: Street markets influencing urban consumer behavior in Mexico. Lat. Am. Bus. Rev. **11**, 77–110 (2010). https://doi.org/10.1080/10978526.2010.487028
3. Train, K.E.: Discrete Choice Methods with Simulation. Cambridge University Press (2009)
4. Nogueira, L.R., de Mello, F.M., de Aguiar, B.S., et al.: Access to street markets and consumption of fruits and vegetables by adolescents living in São Paulo, Brazil. Int. J. Environ. Res. Public Health **15**, 1–12 (2018). https://doi.org/10.3390/ijerph15030517
5. Pitts, S.B.J., Gustafson, A., Wu, Q., et al.: Farmers' market use is associated with fruit and vegetable consumption in diverse southern rural communities 1–11 (2014)

6. Tong, D., Ren, F., Mack, J.: Locating farmers' markets with an incorporation of spatio-temporal variation. Socioecon Plann. Sci. **46**, 149–156 (2012). https://doi.org/10.1016/j.seps.2011.07.002
7. Widener, M.J., Metcalf, S.S., Bar-Yam, Y.: Developing a mobile produce distribution system for low-income urban residents in food deserts. J. Urban Health **89**, 733–745 (2012). https://doi.org/10.1007/s11524-012-9677-7
8. Espejo-Díaz, J., Granados, D., Mejía, G.: Mathematical approach for street markets location in food desert regions: case-study Valparaíso-Chile region. In: Trends in Industrial Engineering Applications to Manufacturing Process. Springer-Verlag
9. López-Ospina, H., Agudelo-Bernal, Á., Reyes-Muñoz, L., et al.: Design of a location and transportation optimization model including quality of service using constrained multinomial logit. Appl. Math. Model. **89**, 428–453 (2021). https://doi.org/10.1016/j.apm.2020.07.054
10. Zambrano-Rey, G., López-Ospina, H., Pérez, J.: Retail store location and pricing within a competitive environment using constrained multinomial logit. Appl. Math. Model. **75**, 521–534 (2019). https://doi.org/10.1016/j.apm.2019.05.040
11. Schön, C., Saini, P.: Market-oriented service network design when demand is sensitive to congestion. Transp. Sci. **52**, 1253–1275 (2018). https://doi.org/10.1287/trsc.2017.0797
12. Zetina, C.A., Contreras, I., Cordeau, J.F.: Profit-oriented fixed-charge network design with elastic demand. Transp. Res. Part B Methodol. **127**, 1–19 (2019). https://doi.org/10.1016/j.trb.2019.06.004
13. Khosravi, S., Akbari Jokar, M.R.: Facility and hub location model based on gravity rule. Comput. Ind. Eng. **109**, 28–38 (2017). https://doi.org/10.1016/j.cie.2017.04.005
14. Alcaldía Mayor de Bogotá D.C.: Decreto 190 de 2004 (2004)
15. Sharma, P., Phanden, R.K.: Analysis of site selection based on factors rating. Int. J. Emerg. Trends Eng. Dev. **6**, 616–621 (2017)
16. Mejía, G., Guerrero, W., Sarmiento, A., et al.: Delivery of perishable export products in smart cities: a case study in Bogotá (Colombia). Proc. Manuf. **39**, 1946–1952 (2019). https://doi.org/10.1016/j.promfg.2020.01.237
17. Fernandez-Anez, V.: Stakeholders approach to smart cities: a survey on smart city definitions BT—smart cities. In: Alba, E., Chicano, F., Luque, G. (eds.), pp. 157–167. Springer International Publishing, Cham
18. Haase, K., Müller, S.: A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. Eur. J. Oper. Res. **232**, 689–691 (2014). https://doi.org/10.1016/j.ejor.2013.08.009
19. National Administrative Department of Statistics (DANE): Proyecciones de población—Bogotá (2018)
20. FAO, Ministerio de Salud y Protección social: Perfil nacional de consumo de frutas y verduras (2013)
21. IDECA: Sistema Distrital de Parques y Escenarios Públicos Deportivos. Bogotá D.C. In: Datos Abiertos Bogotá (2020)
22. (2016) ¿El supermercado o la plaza? Los contrastes de precios de alimentos. El tiempo
23. Acosta Leal, D.A.: Fijación de precios en mercados campesinos de Bogotá. Caso hortalizas frescas de Fómeque y Chipaque (Cundinamarca) (2014)
24. Koroliuk, M., Connaughton, C.: Analysis of big data set of urban traffic data 1–12 (2015)
25. Portafolio: Colombia, el país donde las 'tiendas de barrio' dominan. In: Finanzas|Economía (2013)
26. Logyca: HÁBITOS Y PREFERENCIAS DE LOS CONSUMIDORES DE LAS TIENDAS DE BARRIO (2014)
27. Secretaría de Planeación, Alcaldía Mayor de Bogotá D.C.: Conociendo la localidad de Engativá (2009)

28. IDECA: Alcaldía Mayor de Bogotá D.C. Datos Abiertos Bogotá
29. DANE: Metodología de Estratificación Socioeconómica urbana para servicios públicos domiciliarios (2015)
30. Guzman, L., Arellana, J., Cantillo-Garcia, V.: El estrato socioeconómico como variable sustituta del ingreso en la investigación en transporte. Evaluación para Bogotá, Medellín, Cali y Barranquilla

Sustainable Production

Smart Eco-Factory—Aspects for Next Generation Facilities Supporting Sustainable and High-Tech Production

Michael Hertwig and Joachim Lentes

Abstract Producers are increasingly faced with greater challenges. The requirements in the manufacturing of high-tech products are constantly increasing, especially the demands on the production environment are becoming more relevant. The market demands change with high volatility. In addition, debates and subsequent changes in the laws on sustainable management and resource optimization are increasing, also based on growing public awareness. The concept of smart network production is a promising approach for industrial companies to address the resulting dynamics. Industry 4.0 measures allow production of smaller lots and higher manufacturing flexibility. Connected automation technologies and adaptable assisting technologies support these efforts. However, the flexibilization of manufacturing is only possible if the building offers suitable framework conditions. Factory buildings have a long-lasting life cycle even up to 30 years and more. Therefore, the aspects of optimization of energy consumption and high flexibility of the production areas need to be considered in planning projects, too. Additionally, the support of digital production and smart building maintenance need to be planned in an early stage. To reduce the impact on the surrounding, the integration of factories into the environment as well as the definition of symbiosis with urban infrastructures need to be considered in ultra-efficient production and innovation areas. In the context of the article, it shall be discussed how current trends in digitalization of the production affect the planning and realization of building processes, as well as the question, which requirements need to be considered to realize production infrastructure serving future needs. An important factor to realize energy efficiency and sustainability does not only cope with the building itself. The integration of aspects concerning factory surroundings are also important in supporting acceptance for new constructions and production units. Compliant production and offerings by local enterprises increase the positive social impact on the surroundings. The analysis of a pioneer company in the field of factory building construction enables the evaluation of economic business development combined with sustainable action. The implementation of currently

M. Hertwig (✉) · J. Lentes
Fraunhofer Institute for Industrial Engineering IAO, Nobelstrasse 12, 70569 Stuttgart, Germany
e-mail: michael.hertwig@iao.fraunhofer.de

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_61

existing technology support fulfils the objective, generating a positive impact along the lifecycle of the building.

Keywords Sustainable production · Eco-industrial estates · Future-oriented factory building · Sustainable factory construction · Resource-efficient factory operation

1 Introduction and Motivation

The need to reduce the resource consumption is increasing. A paradigm shift is necessary towards more sustainable production, to stay within the resource limits of the earth. With the increase of wealth of the society and output of production result in arise of resource consumption worldwide. Therefore, the Earth overshot day was terminated early every year [1]. For this purpose, a shift in sustainable proportionality between resource use and production volume is a key element [2]. In particular, this includes optimization of resource consumption combined with increasing effectiveness of enterprises [3]. This is increase of wealth is also correlated with the population living in cities which also impacts the ecological footprint [4]. Until 2050, 70% of the world's population are anticipated to live in cities and even today 75% of the global resource consumption are used by cities [5, 6]. This illustrates the need for a paradigm shift to enable a setup of productions in an urban environment worth living in. The main initiator of these negative environmental effects are businesses, which are still forced to maintain financial goals in addition to the needed improvement of their environmental performance. Thus, holistic approaches are required which ensure an environmentally compatible and at the same time financially viable production.

2 Current Approaches for Sustainable Manufacturing Areas

Enterprise perspective

Fraunhofer has developed the concept of the Ultra-Efficient Factory with a view to the sustainable further development of companies [7]. This postulates completely loss-free production. The concept aims to combine the conventional paradigms of efficiency and effectiveness [8]. Ideally, materials and substances are reused or recycled. The aim is to close cycles for both energy flows and material flows. Regenerative and sustainable energy production complete the concept [9]. To enable a holistic view, five fields of action—material, energy, emission, human/staff and organisation—are defined, which build the singularity of the Ultra-Efficient Factory concept and represent an extended sustainability framework [10]. The objective is to achieve Ultra-Efficiency in all five fields of action [11].

According to the approach of urban production, established symbiotic relationships create the potential to realize a positive contribution to the environment [12].

The Positive-Impact Factory takes up this approach and focuses on interactions with neighbors in the fields of material and energy [13]. Furthermore, it postulates an integration of the factory building in its surrounding environment. The second perspective is the optimization of the factory structure to create positive environmental and social impacts. A suitable building structure is required to provide a liable base. At the same time, current trends of digitalization are taken into account in order to design products sustainably and without losses [14].

Industrial park perspective

Since the further development of companies in accordance with sustainability is limited when focusing solely on internal processes, the analysis must be expanded [15]. The consideration of industrial estates in Europe is a suitable approach. Most industrial companies are located in such categorized areas [16]. Due to the local proximity of the companies, it is easy to identify interactions. Through the cooperation of several local stakeholders, a greater effect can be achieved than by each company alone [17]. A hurdle in the cooperation are often the different goals and success criteria of each company. Due to the different ownership structures, a common site-specific long-term strategy must be supported by most of the local companies [18]. Based on this holistic perspective, the approach of ultra-efficiency was extended to the business parks in order to derive concrete impulses for further development. The approach is addressing the five action fields by establishing a cooperation between the enterprises. Different measures support the reduction of emission and losses or optimizes the resource consumption for all participating parties [19].

3 Advancing of Production Areas

In the new development of production areas, the aspect of sustainability can be anchored in the planning at an early stage. In this way, the potential can be leveraged both in realization and during use. To derive requirements for the further development of industrial buildings, both domain experts and practice partners were interviewed. Aspects of relevance have been identified and transformed in topic areas.

3.1 Topic Areas

The aggregation resulted in the derivation of requirements for the design of future-proof production and logistics buildings. For the collection of relevant topics, current research topics in the context of building design and construction were researched. This examination was supplemented by an exchange with various experts. There were exchange formats with experts in production, digitalization, resilience research, urban planning, architecture and organizational development. This interdisciplinarity enabled the identified trends to be clustered. This made it possible to create a trend

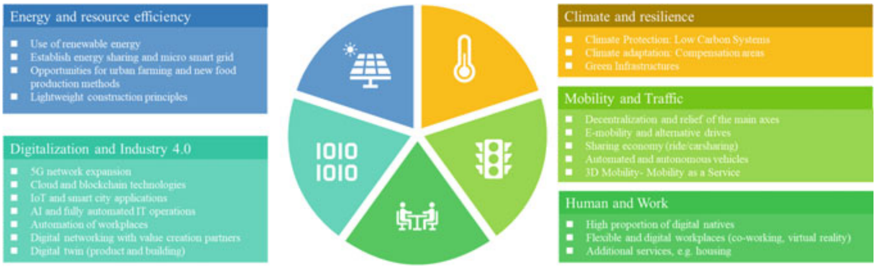


Fig. 1 Identified topic areas of the trend radar

radar with five topic areas (see Fig. 1). These topic areas combine trends of different readiness.

An exchange with industrial users confirmed the identified topic fields. Nevertheless, the interviews revealed additional fields of action. The requirements for the production environment are becoming stricter. Increasing product quality requirements and cost pressure demand stable room conditions in terms of temperature, humidity and air quality. Vibrations must also be absorbed more effectively. Current development cycles are decreasing, especially information and communication technology is developing rapidly. To ensure the usability of the building over its lifetime, the building must be prepared for retrofitting or upgrading. The technical building equipment must be dimensioned accordingly to meet requirements that cannot yet be formulated in concrete terms. This also increasingly includes flexible layout design to meet fast changing production conditions. Increased attention is also being paid to sustainability design and implementation. In this way, potential resource savings can already be directly utilized during realization as well as during use [20]. Resource and energy efficiency potentials can be increased by a long service life, because the emission contribution of construction compared to use increases due to insulation and efficient systems.

3.2 *Climate and Resilience*

Current discussions repeatedly focus on climate change and the associated effects. In order to make a contribution here as a project developer, measures must be identified accordingly. Materials with a low carbon footprint should be used [21]. The extraction and processing as well as the transport aspects should be taken into account. This can already be taken up in the planning, because with the increasing isolation and optimization of the use of energy, the contributions of the construction have a stronger impact [22]. Through the economical use of sealing, compensation areas can be created. These can be greened and planted to reduce the negative input. The realization of green areas positively influences the microclimate and increases the air quality [23]. These approaches can be supported by the realization of green infrastructures, such as green walls and roofs. By improving the air quality and the microclimate, air conditioning and air treatment efforts can be reduced.

3.3 Mobility and Traffic

The mobility sector is currently undergoing a transformation [24]. On the one hand, the switch from combustion engines to alternative drive systems such as fuel cells and batteries is having a major impact. But also new business models based on sharing and renting are leading to jointly used transport units. This is fueled by the discussion of the space requirements of stationary vehicles. In newly developed production areas, approaches can already be integrated at an early stage and thus have a supporting effect. Providing electric mobility infrastructure for automobiles and bicycles increases the acceptance of this transformation [25]. Consideration of public transport infrastructures can also positively influence the use of these modes of transport. In the long term, mobility will be extended into the third dimension. This can already be envisaged as a possible extension [26]. In this way, the attractiveness of the area can be significantly increased or kept high for a longer period of time. A traffic analysis shows the current loads on existing infrastructures. Skillful planning allows the reduction of the load by additional units or allows a relief of the infrastructures [27]. Picking and buffering can be done jointly to optimize delivery and pick-up traffic.

3.4 Human and Work

There is an increasing number of workers available who are referred to as digital natives [28]. This group of workers has already grown up with digital devices and has a high affinity for technology and the use of software to increase efficiency and effectiveness. At the same time, networking and the associated requirements to work together digitally are increasing [29]. Accordingly, especially in product development and virtual collaboration, systems are being worked on that allow people to work together on development data. Based on digital video telephony, tools for virtual collaboration are currently being researched and developed. Thus, location-independent collaboration is to be further supported. The goal is to work together on products in virtual rooms in the long term and to be able to make changes simultaneously in virtual space [30]. The pandemic in particular has shown that a new approach to collaboration is emerging. For example, virtual collaboration tools allow for increasingly flexible work arrangements, such as working in a home office or on the road [31]. The company's premises will increasingly meet specific requirements [32]. For example, specific project rooms for project-related collaboration or innovation areas will become increasingly important. The typical office space will not die out, but the trends of non-territorial working will be increasingly demanded and promoted [33]. This will also influence the education and training of employees [34]. From this it can be deduced that companies must design their office space attractively. The work experience must be improved through the provision of space concepts and technologies. Employees will use the offered office space if the benefit for the work result outweighs the transport effort.

3.5 Digitalization and Industry 4.0

With increasing digitization, companies are increasingly able to react more flexibly to market requirements. This is because the data and information from machines and processes increases transparency [35]. Flexibility can have a positive effect on the utilization of existing machinery. However, the short cycle change of product lines can also lead to changes in layout and organization. In this case, the changeability of the production building will become more important. To realize this, the digital representation must be derived and created in parallel to the realization of the physical building. With methods of Building Information Modeling, a digital twin of the industrial building can be created [36]. If this is linked with the information models of digital production, the adaptation of the building use can be digitally verified at an early stage [37]. This minimizes costs and errors. At the same time, the installed IT infrastructure must be an enabler of digitally supported production. Because only if the data of the machines and production units can be collected and made available, at best in real time, can the desired digitization in production, with the goal of living Industry 4.0 systematically, succeed. In this context, the consideration of requirements of 5G technologies or the installation of 5G-ready devices ensures future orientation [38]. In addition to data networking, this also includes the communication of structural elements such as culverts and gates. Energy storage systems and lighting elements that can be controlled as needed can also be integrated into this networked communication. Thus, the building becomes a smart hub. It supports the seamless flow of processes and real data flows back into the digital representation.

3.6 Energy and Resource Efficiency

The standards of building construction are becoming increasingly stringent. Accordingly, industrial properties are also being provided with better building insulation and insulation [39]. The higher the insulation achieved, the lower the energy required for room temperature control. This can also be furthered by making the best possible use of daylight. A positive contribution can be generated by using renewable energy sources to cover the energy demand [39]. Synergies can be created by creating micro smart grids within the production area [40]. By installing common storage units, excess energy can be buffered and used at later times, e.g. for lighting. Energy sensors can also be installed in the buffers to better compensate for fluctuations over time

[41]. The use of facades and free roof areas for urban farming can also contribute to resource conservation. The impact of sealing is reduced as replacement areas are created. Alternatively, there are early examples of building over existing buildings for improved use of space. Rooftop glasshouses, for example, offer the possibility of producing food close to users in a resource-saving way. Waste heat, for example, can promote plant growth [42]. Furthermore, resource consumption can already be optimized during construction by using lightweight construction principles. New technologies, such as additive technologies, make it possible to manufacture high-strength components using very little material [43]. Their validation and optimization can be further increased by simulation.

4 Analysis of an Exemplary Case

4.1 Introduction of the Pioneer Company

The company under review is a medium-sized enterprise. The company develops and builds logistics and industrial parks with production and storage areas. The usual size of the developments is between 20,000 and 60,000 m^2. Many years of experience both in the logistics sector and in real estate development enable the company to realize individual space concepts. Typical clients are companies from the logistics, retail, manufacturing and service sectors [44].

The company "greenfield development" focuses on high-quality facilities and sustainable construction procedures. This also contains the revitalization of brown-field sites to reduce the consumption of land resources. This maxim is also reflected in project planning, project development and architecture [44].

4.2 Approaches for Long-Term Sustainability for Production Buildings

The focus on long−lasting of commercial real estate lies in two aspects. The company wants to realize compatible buildings. Since some of the properties remain in company ownership and are leased to users, they are designed for resilience, longevity and flexible use. To maintain the high attractiveness of the properties, emphasis is placed on high−quality equipment and solid construction during planning and implementation [45].

Particularly in the technical building equipment, energy−saving approaches are used, such as demand−oriented lighting by means of LED strip lighting, in addition to the use of daylight. Wherever it makes sense, production areas are realized on several floors in order to achieve greater space efficiency. Photovoltaic systems on all roof surfaces with corresponding central energy storage systems are now part of the standard equipment. Where permitted, water reservoirs for sprinkler systems are also used as heat storage. This positive energy concept is complemented by a high level of insulation of the facades and roof surfaces [45].

Furthermore, the equipment features are designed in such a way that expansion over the service life is possible. Thus, appropriate structural facilities are sufficiently dimensioned to ensure this. Fire suppression systems are also provided to make the building suitable for use as a logistics warehouse or as a production area, which increases flexibility [45].

Wherever possible, an attempt is made to embed the building in its surroundings. On the one hand, the color of the facades is adapted to the surroundings, or the terrain structure is used to make the buildings appear less massive.

5 Summary and Outlook

In addition to the efforts of production companies to reduce emissions and optimize material consumption, the reactivation of industrial sites represents great potential for long-term and sustainable use. To leverage this potential, this must be taken into account at an early stage in the development of use. After all, only if the building and the installed infrastructure are future-oriented can a long useful life, high flexibility of use and high attractiveness of the industrial property for potential users be maintained over its lifetime.

This paper shows some aspects that contribute to sustainable production areas over the entire life cycle. However, this paper reflects an approach which need to be evaluated by industrial users. The results of the topic areas need to be implemented in existing buildings and planned construction projects to elaborate the benefits of taking these sustainable measures for industrial and production areas. Together with the frontrunner company the exchange is ongoing. Additional projects are identified where these measures shall be implemented. With the possible evaluation the positive impact can be described.

Acknowledgements We gratefully acknowledge the support of the company "greenfield development" for supporting this project financially and with their perspective.

References

1. Earth Overshoot Day: Past Earth Overshoot Days—#MoveTheDate of Earth Overshoot Day. https://www.overshootday.org/newsroom/past-earth-overshoot-days/ (2021). Accessed 18 May 2021
2. Fischer-Kowalski, M., von Weizsäcker, E.U., Ren, Y., et al.: Decoupling Natural Resource Use and Environmental Impacts from Economic Growth, Paris (2011)
3. Sauer, A., Bauernhansl, T.: Energieeffizienz in Deutschland—eine Metastudie: Analyse und Empfehlungen, 2nd edn. Springer Vieweg, Berlin (2016)

4. Bauer, W.: Wachstumskräfte in den Städten entfesseln: Die Morgenstadt-Initiative. Energy Talks Ossiach 2013, Ossiach, 6 June 2013
5. United Nations: World urbanization prospects, the 2014 revision. Highlights. United Nations, New York (2014)
6. Urban Resource Flows and the Governance of Infrastructure Transitions: www.journals.elsevier.com/environmental-development/news/urbanresource-flows-and-the-governance. Accessed 14 March 2019
7. Mandel, J., Schliessmann, U., Lentes, J., Stender, S.: Die Ultraeffizienzfabrik (UltraEff-UP)—Ressourcenschonende Produktionstechnologien ohne Emissionen im urbanen Umfeld. Forschungsbericht BWPLUS, FKZ L75 13009, Stuttgart (2016)
8. Lentes, J., Mandel, J., Schliessmann, U., Blach, R., Hertwig, M., Kuhlmann, T.: Competitive and sustainable manufacturing by means of ultra-efficient factories in urban surroundings. Int. J. Prod. Res. (2017). https://doi.org/10.1080/00207543.2016.1189106
9. Kuhlmann, T., Sauer, A.: Ultra-efficiency factory—framework and holistic approach to increase and evaluate efficiency and sustainability. Adv. Mater. Res. **11**, 481–488 (2016)
10. Miehe, R., Stender, S., Heßberger, N., Mandel, J., Sauer, A.: Improving manufacturing systems with regard to the concept of ultra-efficiency (2017)
11. Singh, S., Hertwig, M., Lentes, J.: Economic impact of ultraefficient urban manufacturing. Germany—Stuttgart. Part IV. In: Vinod Kumar, T.M. (ed.) Smart Economy in Smart Cities, pp. 273–294. Springer Singapore, Singapore (2017)
12. Lentes, J.: Urbane Produktion. In: Spath, D., Westkämper, W. (eds.) Handbuch Unternehmensorganisation. Springer Reference Technik, Berlin Heidelberg
13. Herrmann, C., Blume, S., Kurle, D., Schmidt, C., Thiede, S.: The positive impact factory-transition from eco-efficiency to eco–effectiveness strategies in manufacturing. Proc. CIRP (2015). https://doi.org/10.1016/j.procir.2015.02.066
14. Herrmann, C., Schmidt, C., Kurle, D., Blume, S., Thiede, S.: Sustainability in manufacturing and factories of the future. Int. J. Precis. Eng. Manuf. Green Tech. (2014). https://doi.org/10.1007/s40684-014-0034-z
15. Pichlmeier, F.: Ressourceneffizienzpotenziale von Gewerbegebieten. VDI Zentrum Ressourceneffizienz GmbH (VDI ZRE), Berlin (2018)
16. Funk, M., Leuninger, S.: Das Gewerbegebiet der Zukunft—Wettbewerbsfähige Standortentwicklung für Kommune durch strategisch-nachhaltige Flächenentwicklung (SNG). Standort **2014**, 153–159 (2014)
17. Jacobsen, N.B.: Industrial symbiosis in Kalundborg, Denmark—a quantitative assessment of economic and environmental aspects. Massachusetts Institute of Technology and Yale University. J. Ind. Ecol. **10**(1–2) (2006)
18. Spath, D., Hertwig, M.: Holistische Standortentwicklung von produzierenden Unternehmen unter Berücksichtigung von Wechselwirkungen mit dem Umfeld. Konzeptstudie. University of Stuttgart, Stuttgart (2016)
19. Hertwig, M., Bogdanov, I., Beckett, M., Waltersmann, W., Lentes, J.: Symbiotic loss-free industrial production in ultra-efficient urban industrial parks. In: 28th CIRP Conference on Life Cycle Engineering. Procedia CIRP (2021). https://doi.org/10.1016/j.procir.2021.01.167
20. Zimmermann, N.: Potentiale durch Additive Technologie. Anwendungsbereiche bei der Gebäudeausrüstung. 18. Fenster-Türen-Treff, Alpbach, 8 March 2019
21. UNDP and Climate Change, New York (2016)
22. Hirsch, T., Lottje, C., Netzer, N.: Exploring sustainable low carbon development pathways. In: Pioneers of Change 21 Good Practices for Sustainable Low Carbon Development in Developing Countries, Berlin (2015)
23. Battisti, A., Santucci, D.: Activating Public Space. An Approach for Climate Change Mitigation. Technische Universität München, Munich (2020)
24. Simpson, C., Ataii, E., Kemp, E., Zhang, Y.: Mobility 2030: transforming the mobility landscape. In: KPMG International Cooperative, Quincy MA (2019)
25. The Evolution of Mobility, Munich (2018)

26. Duwe, D., Keicher, L., Ruess, P., Klausmann, F.: QUO VADIS 3D mobility. Technological readiness, urban and rural use cases & urban integration of flying cars and passenger drones. Fraunhofer IAO, Stuttgart. http://publica.fraunhofer.de/dokumente/N-569614.html (2019)
27. O'Connor, A., O'Brien, E.J.: Traffic load modelling and factors influencing the accuracy of predicted extremes. Can. J. Civ. Eng. (2005). https://doi.org/10.1139/l04-092
28. Puybaraud, M.: Digital Natives: a tech-savvy generation enters the workplace. Expert insight. https://www.workdesign.com/2012/02/digital-natives-a-tech-savvy-generation-enters-the-workplace/ (2012). Accessed 20 May 2021
29. Judd, T.: The rise and fall (?) of the digital natives. AJET (2018). https://doi.org/10.14742/ajet.3821
30. Bues, M., Wingert, B., Riedel, O.: VD1: a technical approach to a hybrid 2D and 3D desktop environment. In: IEEE 11th Workshop on Software Engineering and Architectures for Real-time Interactive Systems (SEARIS), vol. 11, pp. 1–7 (2018)
31. Stahl, A.: The future of offices and workspaces, post-pandemic. https://www.forbes.com/sites/ashleystahl/2021/04/16/the-future-of-offices-and-workspaces-post-pandemic/?sh=1aedc1986442 (2021). Accessed 21 May 2021
32. Pladson, K.: How coronavirus is changing our office spaces. https://p.dw.com/p/3os6c (2021). Accessed 21 May 2021
33. Czycholl-Hoch, H.: The new normal: we still need offices, but not as we know them. https://www.dreso.com/de/en/dossier-the-workplace-reloaded/the-new-normal-we-still-need-offices-but-not-as-we-know-them (2021). Accessed 21 May 2021
34. Schultze, T., Bues, M.: Evaluation digitaler Aus Und Weiterbildung Im Virtuellen Raum. In: Thomas, O., Metzger, D. (eds.) Digitalisierung in der Aus-und Weiterbildung: Virtual und Augmented Reality für Industrie 4.0, pp. 113–123. Springer Gabler, Berlin (2018)
35. Gärtner, S.: Strukturwandel und Produktionsarbeit im urbanen Raum. Arbeit (2019). https://doi.org/10.1515/arbeit-2019-0018
36. Enabling an Ecosystem of Digital Twins (2020)
37. Sacks, R., Brilakis, I., Pikas, E., Xie, H.S., Girolami, M.: Construction with digital twin information systems. DCE (2020). https://doi.org/10.1017/dce.2020.16
38. O'Halloran, J.: In-building wireless infrastructure, 5G indoor revenues will exceed $16bn by 2025. https://www.computerweekly.com/news/252491161/In-building-wireless-infrastructure-5G-indoor-revenues-will-exceed-16bn-by-2025 (2020). Accessed 25 May 2021
39. Energy efficiency in buildings. Module 18, Wien (2016)
40. Guan, X., Xu, Z., Jia, Q.: Energy efficient buildings facilitated by micro grid. IEEE Trans. Smart Grid (2010). https://doi.org/10.1109/PES.2010.5590040
41. Loschi, H.J., Leon, J., Iano, Y., Filho, E.R., Conte, F.D., Lustosa, T.C., Freitas, P.O.: Energy efficiency in smart grid: a prospective study on energy management systems. SGRE (2015). https://doi.org/10.4236/sgre.2015.68021
42. Reisinger, M., Sauer, A.: Urban production: smart rooftop greenhouses on factories. ECEEE Summer Study Proc. (2017). https://doi.org/10.1016/j.jclepro.2011.04.016
43. Mahler, B., Idler, S., Nusser, T., Gantner, J.: Energieaufwand für Gebäudekonzepte im gesamten Lebenszyklus. TEXTE, 132, Dessau-Roßlau (2019)
44. Greenfield development: Website of enterprise. https://greenfield-development.de/ (2021). Accessed 18 May 2021
45. Kuehne, C.: Commercial Buildings Developed by Greenfield Development. Waiblingen (2020)

A Green Approach on Multiple Allocation Hub Covering Flow Problem

Nazmi Sener, Hasan Hüseyin Turan, Fuat Kosanoglu, and Mahir Atmis

Abstract Green supply chain applications are gaining wide interest as environmental and economic concerns have been increasing. Transportation authorities impose strict regulations to reduce the negative effect of transportation activities on the environment such as carbon emissions and noise. In this study, we introduce a novel optimization model for a green multiple allocation hub covering problem that seeks to find the best locations for hubs and allocation of origin/destination flows in a logistic network. Different from the existing hub covering flow problems, carbon emission costs is added to the total cost function aiming to minimize the total amount of emissions in the network. Further, we investigate the effect of several factors (e.g., coverage radii of hubs) on optimal solutions. To illustrate the efficiency of our approach over the conventional hub location models, we first generate a new dataset by combining a well-known dataset used in traditional hub covering problems with a carbon emission dataset provided by the International Air Transport Association (IATA). Afterward, we present a comparison of our model to previous models that do not consider carbon emission.

Keywords Green supply chain · Hub covering · Carbon emission · Optimization

1 Introduction

The design of transportation network plays an important role in developing a sustainable supply chain network [1]. Transportation sector is considered among the leading sectors with harmful effects on environment, including excessive natural resource

N. Sener (✉)
Industrial Engineering Department, Alanya Alaaddin Keykubat University, Antalya, Turkey
e-mail: nazmi.sener@alanya.edu.tr

H. H. Turan
Capability Systems Centre, University of New South Wales, Canberra, Australia

F. Kosanoglu · M. Atmis
Department of Industrial Engineering, Yalova University, Yalova, Turkey

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_62

usage, acidification, land use, adverse effects on wildlife, noise, and the Greenhouse Gas (GHG) emission. GHG emissions, especially CO_2 emissions, and noise pollution are the most disturbing effects, since they directly affect human health [2]. Moreover, increasing amount of GHG emissions is one of the main driver to climate change for most countries, especially industrializing nations such as the United Kingdom and the United States [3, 4]. Therefore, increasing concerns on the harmful effects of transportation activities to human health impose more effective and eco-friendly transportation network planning [5].

The hub location problems lie at the heart of network design planning in many industries such as transportation [6], telecommunication [7], emergency services [8], supply chain management and logistics [9], and computer systems [10]. The main concern of these models is locating hub nodes and assigning non-hub nodes (a.k.a. spokes) to the hubs to achieve a certain objective (such as minimizing total transportation and operation cost in the hub network). Campbell and O'Kelly [11] classify HLP according to similarity of objective functions to well-known facility placement problems as hub median, hub center, and hub covering location problems. The problem is a vertex location problem, since hub locations are limited to network nodes [12].

Most of HLPs are based on two fundamental assumptions: (1) flows must be routed through a collection of nodes that includes at least one hub node, and (2) the HLP is a network problem with more efficient and higher volume paths that enable a discount factor α $(0 < \alpha < 1)$ to be implemented to all distribution costs associated with flows transported among each couple of hub nodes [13].

In this paper, we focus on hub covering problem (HCP) which is a type of HLP. HCP is analogous to set covering problem in application of HLP. There are a limited number of studies in the literature addressing hub covering problems. Campbell [14] studied the first hub covering problem. In this study, hub set cover problem and maximal hub cover problems were introduced with three different coverage scenarios for single and multiple allocations cases. In particular, in the first coverage scenario, hubs k and$l$cover origin/destination (O-D) pairs (i, j), if the distance of i to j via hubs k and/or$l$does not exceed a predetermined value. In the second coverage scenario, hubs k and$l$cover O-D pairs (i, j) if the distance of each link in the path of i to j via hubs k and/or$l$does not exceed a specified value. The third coverage scenario assumes that hubs k and$l$cover O-D pairs (i, j) if each of hub-origin/destination distance within specified range.

Kara and Tansel [15] propose a new linearized formulation of HCP. Kara and Tansel [15] model's outperforms to the best performing linearization of Campbell's model [14] in terms of average CPU times and storage requirements. Tan and Kara [16] develop a hub infrastructure model with an application to a Turkish cargo company. Wagner [17] suggests an updated formulation to hub covering problem that addresses both single and multiple allocation issues. Alumur and Kara [18] implement an extension of Tan and Kara [16] in cargo company. Calik et al. [19] present an integer programming model for solving the hub covering problem with a single allocation over an incomplete network. Lowe and Sim [20] develop a new single allocation hub covering problem formulation that includes transportation costs. Karimi et al. [21] present a model that incorporates multimodal concept with the capacitated single allocation p-hub covering problem. Sener and Feyzioglu [13] enhance the multiple allocation hub covering formulations that are proposed by Lowe and Sim [20].

Mohammadi et al. [5] study the first green concerned hub location problem. Niknamfar and Niaki [22] develop a novel bi-objective green hub location problem with the centralized carrier collaboration framework for multiple carriers where one objective is for carbon emission and one objective for transportation cost. Dukkanci et al. [23] introduce the first single objective green hub location problem and solve with second order cone programming method. Parsa et al. [24] develop a multi objective hub location problem that aims to minimize greenhouse emission, noise pollution, and fuel consumption.

We present relevant literature review in Table 1. As shown in Table 1, there are limited number of studies on hub covering problems. Best of our knowledge, there are several studies on green approaches to hub location problem, yet none of them is hub covering problem. This study is an initial attempt to apply green approaches in hub covering location problem. We organize the rest of the paper as follows as follows. Section 2 of this paper presents mathematical model. Numerical results are discussed in Sect. 3. Finally, the last section discusses the conclusions from our work and gives directions for possible future research.

2 Mathematical Model

In this section, we present a novel mathematical model for multiple allocation hub covering flow problem with carbon emission costs (MAHCFP-CE). The notation of the model is given below.

i, j	node indices
k, l	hub indices
f_k	the fixed cost of establishing a hub facility in node k
c_{ij}	the unit transfer cost from origin node i to the destination node j
h_{ij}	total demand transferred from node i to node j
O_i	total demand sent from node i
D_j	total demand delivered to node j
d_{ij}	distance from node i to node j
α	interhub transportation cost factor ($0 < \alpha < 1$)
A_{ik}	node covering indicator (1 if node i can be covered by hub k and 0 otherwise)
B_{ikj}	path covering indicator ($B_{ikj} = A_{ik} A_{kj}$)
ρ	carbon emission cost per CO_2 kg emitted
et_{ij}	carbon emission for transporting a unit from node i to node j (kg/ton)

(continued)

(continued)

ets_{ikj}	carbon emission for transporting a unit from node i to node j through hub k ($ets_{ikj} = et_{ik} + et_{kj}$)
Decision Variables	
$x_k =$	$\begin{cases} 1, & if\ a\ hub\ located\ at\ k \\ 0, & else \end{cases}$
y_{ikl}	amount of the flow which originated from node i to delivered hub l passing through hub k
z_{ik}	amount of the flow which originated from node i to delivered hub k
q_{ilj}	amount of the flow which originated from node i to delivered node j passing through hub k

The resulting mathematical model is given below.

$$\min \sum_k f_k x_k + \alpha \sum_{i,k,l} c_{kl} B_{ikl} y_{ikl} + \sum_{i,k} c_{ik} A_{ik} z_{ik} + \sum_{i,l,j} c_{lj} B_{ilj} q_{ilj}$$
$$+\rho \left(\sum_{i,k,l} ets_{ikl} B_{ikl} y_{ikl} + \sum_{i,k} et_{ik} A_{ik} z_{ik} + \sum_{i,l,j} ets_{ilj} B_{ilj} q_{ilj} \right) \tag{1}$$

$$\text{s.t.} \sum_k A_{ik} z_{ik} = O_i \quad \forall i, \tag{2}$$

$$\sum_l B_{ilj} q_{ilj} = h_{ij} \quad \forall i, j, \tag{3}$$

$$\sum_l B_{ikl} y_{ikl} + \sum_j B_{ikj} q_{ikj} - \sum_l B_{ilk} y_{ilk} = A_{ik} z_{ik} \quad \forall i, k, \tag{4}$$

$$\sum_i B_{ilj} q_{ilj} \le D_j x_l \quad \forall l, j, \tag{5}$$

$$A_{ik} z_{ik} \le O_i x_k \quad \forall i, k, \tag{6}$$

$$q_{ilj}, y_{ikl}, z_{ik} \ge 0 \quad x_k \in \{0, 1\} \quad \forall i, k, l, j. \tag{7}$$

The objective function (1) is to minimize the total cost of establishing hubs, plus demand delivery cost through links by considering interhub transportation cost discount factor, and carbon emission penalty cost sourced from the transportation. Equation (2) ensures that all the demand from node i is delivered through hubs. Equation (3) satisfies that the demand from node i to node j is sent through hubs. Equation (4) guarantees that there is no loss demand in the hubs. Equations (5) and (6) together ensure that a demand is not sent directly from a non-hub node to another non-hub nodes. Finally, Eq. (7) shows the type of decision variables.

Table 1 Summary of the relevant literature

Study	Cov.	Med.	Assignment		Cost		Green
			SA	MA	HC	TC	
Campbell [14]	×	×	×		×		
Kara and Tansel [15]	×		×		×		
Tan and Kara [16]	×		×		×		
Wagner [17]	×		×	×	×		
Alumur and Kara [18]	×		×		×		
Calik et al. [19]	×		×		×		
Lowe and Sim [20]	×		×		×	×	
Karimi et al. [21]	×		×		×	×	
Sener and Feyzioglu [13]	×			×	×	×	
Mohammadi et al. [5]		×	×		×	×	×
Niknamfar and Niaki [22]		×	×			×	×
Dukkanci et al. [23]		×	×			×	×
Parsa et al. [24]		×		×	×	×	×
This paper	×			×	×	×	×

Cov covering, *Med* median, *SA* single allocation, *MA* multiple allocation, *HC* hub establishment cost, *FC* transfer costs

3 Numerical Experiments

We use an extended version of the well-known and openly available CAB dataset [25] in our numerical study. This dataset is consisting demands and locations of 25 United States airports. Furthermore, we use carbon emission calculations which is proposed by IATA [26]. We assume that unit transportation costs are proportional to link lengths such that $c_{ij} = d_{ij}/25{,}000$ for all i, j, and hub establishment costs f_i for all i are set all equal to 10,000, 30,000, and 50,000 for a given instance. Interhub transportation cost discount factor α is chosen as 0.2, 0.5 or 0.8. The hub (node) coverage radius is calculated by multiplying the coverage ratio Δ with the length of the longest link of the network. Δ should be selected such that no disconnected sub-networks are available the network. Therefore, Δ is set to 0.6, 0.7 or 0.8. Then, each element of A_{ij} is assigned to 1 if the node coverage radius is greater than or equal to the length of the link connecting nodes i and j, and 0 otherwise. ρ values are based on carbon pricing values of California, Massachusetts, and Switzerland from the World Bank Carbon Pricing Dashboard which are 16.89, 8.26, 104.65 consecutively.

To demonstrate the effect of the carbon emission on the total cost of proposed MAHCFP-CE, we introduce two indicators: the number of established hubs (NH), and the percent of carbon emission decrease (CD). The carbon emission amount of a MAHCFP-CE model is denoted as $\pi^{\star}_{MAHCFP-CE}$, and the value of (emission) function in Eq. (1) calculated by using the optimum solution of the associated MAHCFP model, which is proposed by Sener and Feioglu [13], is denoted as π_{MAHCP}. CD is calculated as follows:

$$\mathrm{CD} = 100 \times \frac{\pi_{MAHCFP} - \pi^{\star}_{MAHCFP-CE}}{\pi_{MAHCFP}}$$

All instances have been solved with Gurobi 9.1 on dual Intel Xeon E5-2670 (2.6 GHz) processor and 32 GB of RAM workstation running Windows Server 2012 R2-64 bits. Results of the numerical experiments are summarized in Table 2. Carbon emission decrease percentage (CD) boosts as ρ increases, and CD reaches its peak values at the greatest ρ value. Indeed, the change of ρ value sharply diversifies both NH and CD. The hub establishment cost, f_i, has adverse effect on CD. As the total hub establishment cost becomes substantial compared to the total transportation cost, rising hub launching costs obviously reduces NH. The decrease in α leads to more flows between hubs, which encourages the establishment of new hubs to take advantage of the cost savings. However, the effect of α and Δ on CD depends on the other parameters. In general, adding carbon emission cost to the model decreases the carbon emission compared to the MAHCFP model as expected. For example, at least 4% less emission is observed as shown in Table 2. On the other hand, it causes opening more hubs, which results in more complex network structure. The carbon emission costs are dependent to the flow amount as expected.

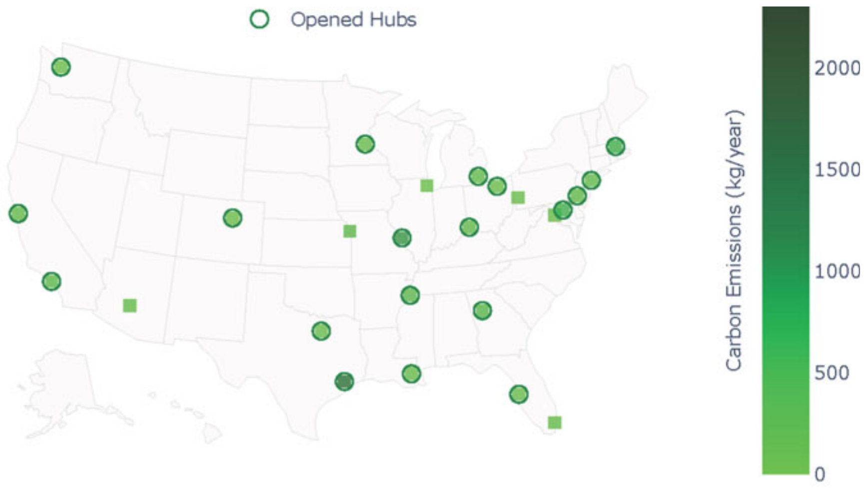
The actual positions of opened hubs and carbon emission amounts of each node for two distinct cases are mapped on the layouts in Fig. 1. Figure 1a is selected because the largest CD is observed in this case. Contrarily, Fig. 1b is chosen because the smallest CD is acquired in this case. It is obvious that hub positions are determined by high accessibility and total demand streaming in/out of a node. Furthermore, cost savings from interhub transfers facilitate the establishment of multiple hubs. Carbon emission amounts of nodes decrease when carbon emission cost increases as expected.

4 Research Insights

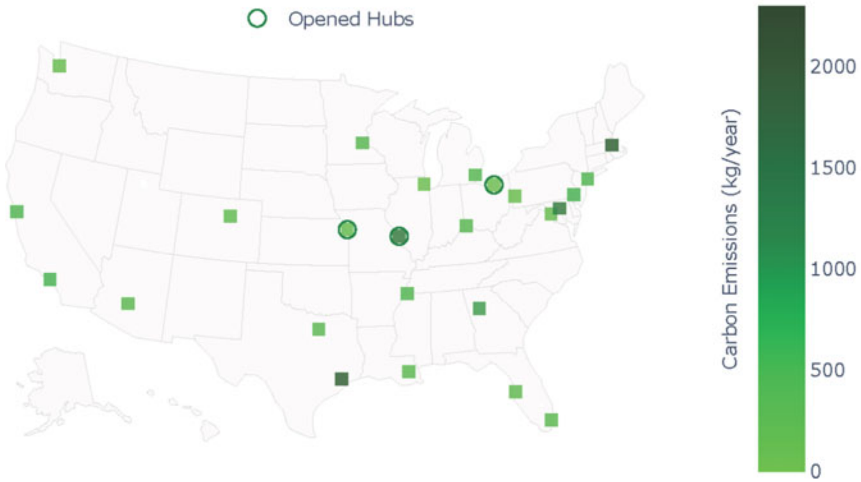
In this section, we derive several research insights from the numerical experiments. First, unlike the conventional hub location problem, increasing the number of hubs to be opened in the MAHCFP-CE does not automatically decrease the objective function value or the total amount of emissions. As a result, there is no certainty that providing a service with a large number of hubs reduces emissions. Second, carbon costs significantly change the network structure and number of opened hubs. Thus, regulations must be taken into account when the network setup. Third, choosing the right coverage radius is very important to reduce emissions.

Table 2 Summary of performance measures

ρ		16.89						8.26						104.65					
Δ		0.8		0.7		0.6		0.8		0.7		0.6		0.8		0.7		0.6	
f_i	α	NH	CD	NH	CD	NH	CD	NH	CD	NH	CD	NH	CD	NH	CD	NH	CD	NH	CD
10,000	0.8	8	31.8	7	32.8	7	27.6	6	18.4	6	22.5	7	20.9	19	45.1	19	46.6	18	41.8
	0.5	8	24.9	8	26.1	8	24.4	8	10.1	8	10.5	7	11.0	19	48.5	19	47.1	19	43.6
	0.2	9	13.1	9	14.9	10	15.6	9	6.5	9	6.8	10	8.3	19	50.0	18	48.0	19	45.9
30,000	0.8	4	20.0	4	21.4	4	22.5	3	9.1	3	11.7	4	21.4	10	37.3	10	37.7	9	37.5
	0.5	4	25.8	4	17.1	4	15.8	4	18.4	4	9.8	4	9.2	10	46.0	10	38.3	9	34.4
	0.2	5	14.7	5	15.1	5	17.0	5	9.3	5	9.7	5	11.0	10	45.0	10	43.5	10	41.8
50,000	0.8	3	11.8	3	13.4	3	16.1	3	9.1	3	11.7	3	13.6	6	29.2	6	32.5	7	35.9
	0.5	3	16.3	3	15.4	3	8.0	3	10.3	3	10.2	3	4.3	6	35.5	6	36.0	7	31.7
	0.2	3	11.8	3	13.4	3	16.1	3	9.1	3	11.7	3	13.6	6	29.2	7	32.5	7	35.9



(a) ρ=104.65, α=0.2, Δ=0.8, f_i=10,000



(b) ρ=8.26, α=0.5, Δ=0.6, f_i=50,000

Fig. 1 Carbon emission amounts of nodes and location of hubs

5 Conclusion

In this study, we aim to find optimal structure of the origin–destination network by taking into account carbon emission, hub establishing, flow costs, and coverage constraints. The demand for a particular origin–destination node pair must be routed through at least one hub node. In this work, it is assumed that a hub covers a nonhub node if the distance between these two is less than a predefined value, while the distance between hubs has no limit.

The numerical results show that the number of hubs is strongly depending on carbon emission cost and hub establishment cost. Furthermore, when the number of hubs opened is relatively small due to carbon emission cost and hub establishment cost, hubs optimally located near the nodes with higher demand.

Our computational study also shows that carbon emission costs can have a significant impact on green network design. For example, total carbon emission of the MAHCFP-CE could be lowered as 50% below compared to MAHCFP as shown as Table 2.

MAHCFP-CE model can be extended in several ways. For example, it might be worthwhile to add different emissions such as black carbon, or sulphuryl. Including link or hub capacity, addition to MAHCFP-CE could be another important extension.

References

1. Elhedhli, S., Merrick, R.: Green supply chain network design to reduce carbon emissions. Transp. Res. Part D Transp. Environ. **17**, 370 (2012)
2. Regional Branch of Europe, W.: Achieving Health Benefits From Carbon Reductions (2018)
3. Department of Transportation UK: Low Carbon Transport: A Greener Future (2009)
4. OECD/ITF: Greenhouse Gas Reduction Strategies in the Transport Sector: Preliminary Report
5. Mohammadi, M., Torabi, S.A., Tavakkoli-Moghaddam, R.: Sustainable hub location under mixed uncertainty. Transp. Res. Part E Logist. Transp. Rev. **62**, 89 (2014)
6. Lin, C.-C., Lin, J.-Y., Chen, Y.-C.: The capacitated p-hub median problem with integral constraints: an application to a Chinese air cargo network. Appl. Math. Model. **36**, 2777 (2012)
7. Kim, H., O'Kelly, M.E.: Reliable p-hub location problems in telecommunication networks. Geogr. Anal. **41**, 283 (2009)
8. Berman, O., Drezner, Z., Wesolowsky, G.O.: The transfer point location problem. Eur. J. Oper. Res. **179**, 978 (2007)
9. Ishfaq, R., Sox, C.R.: Design of intermodal logistics networks with hub delays. Eur. J. Oper. Res. **220**, 629 (2012)
10. Hartoonian, A., Khademzadeh, A.: Hub location allocation problem in computer networks using intelligent optimization algorithms. Int. J. Inf. Commun. Technol. Res. **9** (2017)
11. Campbell, J.F., O'Kelly, M.E.: Twenty-five years of hub location research. Transp. Sci. **46**, 153 (2012)
12. Owen, S.H., Daskin, M.S.: Strategic facility location: a review. Eur. J. Oper. Res. **111**, 423 (1998)
13. Sener, N., Feyzioglu, O.: Capacitated multiple allocation hub covering flow problem. Manas. J. Eng. (2021) (In press)
14. Campbell, J.F.: Integer programming formulations of discrete hub location problems. Eur. J. Oper. Res. **72**, 387 (1994)
15. Kara, B.Y., Tansel, B.C.: The single-assignment hub covering problem: models and linearizations. J. Oper. Res. Soc. **54**, 59 (2003)
16. Tan, P.Z., Kara, B.Y.: A hub covering model for cargo delivery systems. Networks **49**, 28 (2007)
17. Wagner, B.: Model formulations for hub covering problems. J. Oper. Res. Soc. **59**, 932 (2008)
18. Alumur, S., Kara, B.Y.: A hub covering network design problem for cargo applications in Turkey. J. Oper. Res. Soc. **60**, 1349 (2009)
19. Calik, H., Alumur, S.A., Kara, B.Y., Karasan, O.E.: A Tabu-search based heuristic for the hub covering problem over incomplete hub networks. Comput. Oper. Res. **36**, 3088 (2009)

20. Lowe, T.J., Sim, T.: The hub covering flow problem. J. Oper. Res. Soc. **64**, 973 (2013)
21. Karimi, H., Bashiri, M., Nickel, S.: Capacitated single allocation P-hub covering problem in multi-modal network using Tabu search. Int. J. Eng. **29** (2016)
22. Niknamfar, A.H., Niaki, S.T.A.: Fair profit contract for a carrier collaboration framework in a green hub network under soft time-windows: dual lexicographic max-min approach. Transp. Res. Part E Logist. Transp. Rev. **91**, 129 (2016)
23. Dukkanci, O., Peker, M., Kara, B.Y.: Green hub location problem. Transp. Res. Part E Logist. Transp. Rev. **125**, 116 (2019)
24. Parsa, M., Nookabadi, A.S., Flapper, S.D., Atan, Z.: Green hub-and-spoke network design for aviation industry. J. Clean. Prod. **229**, 1377 (2019)
25. Beasley, J.E.: OR-Library. http://people.brunel.ac.uk/~mastjjb/jeb/info.html
26. IATA: Air Cargo Carbon Footprint. https://www.iata.org/en/programs/cargo/sustainability/carbon-footprint/

A Linguistic MCDM Approach to Overcome Future Challenges of Vertical Farming

Deniz Uztürk and Gülçin Büyüközkan

Abstract Food production is an essential operation where production and resource efficiency are low compared to other sectors. With the unstoppable growth of the world population, agricultural production is under pressure to meet the increasing food demand. Controlled Environment Agriculture (CEA) is a successful solution to create sustainable and resilient development through sustainable cities. CEA, where the farming activities are isolated from the meteorological conditions, is one of the most powerful solutions to adapt and mitigate climate change in urban areas. Vertical farming (VF) is also an indoor plant manufacturing process. In VF, plants are grown in layers and can thus reach high. The system can entirely be designed without any dependence on sunlight or other outdoor resources. However, there are a significant number of drawbacks about VF in the literature, such as limited products and labor costs, etc. This study focused on generating the VF area's main challenges and wanted to create a roadmap to overcome these challenges. Existing VF challenges are gathered from experts and related literature. Possible solutions to overcome these limitations are derived from the literature as well. The process is approached as a multi-criteria decision-making (MCDM) procedure. The House of Quality (HoQ) of Quality Function Deployment (QFD) is suggested to investigate the relationships between solutions and challenges. The HoQ method also allows for prioritizing the potential solutions to generate a roadmap for practitioners. Plus, the methodology extends the QFD model with the 2-tuple linguistic model to overcome the vagueness by supplying linguistic sets to decision-makers (DMs) to assess via semantics closer to the human cognitive process. That helps to improve the accuracy of the linguistic computations and interpretability of the results. Also, it creates a flexible environment for DMs. A case study is applied for Turkey, and sensitivity analyses are presented to test the suggested methodology's robustness.

D. Uztürk
Department of Business and Administration, Galatasaray University, Istanbul, Turkey

G. Büyüközkan (✉)
Department of Industrial Engineering, Galatasaray University, Istanbul, Turkey
e-mail: gbuyukozkan@gsu.edu.tr

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_63

Keywords 2-tuple linguistic model · Controlled environment agriculture · MCDM · House of quality · Vertical farming

1 Introduction

The revolution in production systems with Industry 4.0 technologies reaches out to various new sectors by integrating these technologies into conventional systems. The agriculture sector is also affected by this change, and today we are talking about producing food inside the urban buildings. The "Plant factory" notion has emerged from the intersection of Industry 4.0 technologies and traditional agricultural operations [1].

Even the plant factories enable Controlled Environment Agriculture (CEA) and allow food safety for the urban population; they have some drawbacks, as mentioned in the literature. Vertical farming (VF), an approach for indoor agriculture, stands as a potent solution for urban areas. However, some challenges are needed to be overcome to reach a durable urban farming system [2].

Consequently, this paper aims to create a decision-making model to analyze current challenges for better future VF implications. This paper also aims to address future researchers or practitioners by creating a road map for handling the challenges of VF for sustainable food production.

The proposed methodology is based on multi-criteria decision-making (MCDM). The well-known method, House of Quality (HoQ) of Quality Function Deployment (QFD), is recommended to assess the relations between challenges and solutions.

The proposed methodology initially identifies the possible difficulties for VF as customer needs (CNs) and identifies potential solutions to these challenges as design requirements (DRs) in HoQ. The suggested model also uses a group decision-making (GDM) approach to reveal objective expert knowledge about the subject. A linguistic framework is proposed by integrating the 2-tuple linguistic model with HoQ to create a flexible and interpretable decision-making environment for the decision-makers (DMs).

The organization of this paper is as follows: The following section presents the theoretical background by giving explanations about CEA, VF, and VF challenges. Section 3 provides the methodological background by introducing the basics of the 2-tuple model and HoQ. Finally, the case study is given in Sect. 4, and the results, analysis, and conclusions are provided followingly.

2 Theoretical Background

In this section, a piece of brief background information about the application area will be provided. First, CEA's general concept will be presented; later, the VF notion and its challenges generated by the literature and expert views will be given.

2.1 Controlled Environment Agriculture

CEA is defined as "the production of plants and their products, such as vegetables and flowers, inside controlled environment structures such as greenhouses, vertical farms and growth chambers" by the University of Arizona's Controlled Environment Agriculture Center [3]. CEA enables food production with high efficiency and safety; accordingly, it is a valuable solution for future agrarian operations. Today CEA presents a small number of agricultural practices, yet it is still an emerging area with new technological integrations.

The increasing urban population, poor management practices in existing agricultural lands, non-efficient use of spaces in urban areas are the major issues that let the researchers search for more efficient and effective approaches for agriculture [4]. The CEA approach is very suitable for urban areas since it only needs a closed building to produce. In addition to daylight, artificial lighting with different wavelengths also gives the possibility to grow food inside the buildings. These new technologies such as LED lightings, smart building systems, IoT allow us to transform existing buildings into plant production factories [5]. Despite its benefits mentioned above, the CEA approach, like VF, has some drawbacks and limitations for future applications. These drawbacks are similar to the VF's challenges and vulnerabilities, so in the next section, after giving brief information about VF systems, the challenges and solutions will be delivered in Sect. 2.3.

2.2 Vertical Farming

Locally produced vegetables can diminish the food mile, shortening the supply chain [6]. That creates a sustainable system where urban residents easily reach locally produced products. Therefore, urban farming is one of the critical solutions for food security in modern cities. Urban agriculture/farming is the solution, but it cannot exist horizontally in cities, so VF is proposed for the towns [7]. VF is a way of aliment production by using vertically aligned surfaces. The system can be merged with other structures like skyscrapers, warehouses, etc. Reducing the vulnerabilities caused by weather, augmenting the novel job chances, secure food with fewer chemicals, less resource usage are the critical benefits for VF [8]. There will be less transportation thanks to the locally produced foods, preventing spoilage due to excessive handling. In the VF area, there exist three different options to create a VF environment. These are Hydroponics, Aquaponics, Aeroponics [9]. Three of them are the consequences of developed technologies in soilless agriculture. Those agricultural farming systems can be robust solutions to provide different products that require less water, less fertilizer, and less space. Also, they can expand the yield per unit area.

Thanks to emerging technology and R&D researches, the cost of VF is decreasing. Especially in cities, the more strategy and planning are addressed, there will be more straightforward adaptation and shifting to CEA. Plus, educating government officials and farmers to accustom them to new technologies and supporting infrastructure development will be an excellent strategy for future cities [4]. However, despite the developing technologies, there still exist some limitations to address in order to reach more sustainable and durable systems in the future [10].

2.3 VF Challenges and Solutions

In this section, VF's main challenges, vulnerabilities, and drawbacks are collected and compiled from the existing academic and industrial literature. Afterward, meetings are organized with experts, which will be mentioned with details in the case study, and the challenges are rearranged and then validated by the expert group. Table 1 below presents the detected vulnerabilities grouped under social, economic, political, and environmental dimensions [4, 6, 11–16].

Five main dimensions are generated for the VF challenges, and the sub-dimensions are grouped accordingly. As it can be interpreted from Table 1, the most numbered dimension is the cost-related one. Therefore, it is easy to summarize that the most focused challenge is the "cost" dimension of applied technologies and lighting systems in VF [11, 17].

Furthermore, potential solutions to overcome the challenges are also compiled from the academic and industrial literature. The solutions are also validated by the experts, and they are given in Table 2 [4, 12, 14–16].

The main aim of this paper is to prioritize collected solutions by assessing their relations with collected challenges. The HoQ methodology will let the relation assessment by its matrix structure. The details of the suggested methodology will be given in the next section.

Table 1 VF challenges [4, 12, 14–16]

Main dimensions	Sub-criteria
Social	Narrow product range (C1) High selling price (C2)
Economic	High building cost (C3) LED energy use (C4) Competition with greenhouses (C5) Selling difficulty (C6)
Environmental	Visual pollution (C7)
Political	Urban development planning (C8) Lobbying (C9)

Table 2 Potential solutions for future VF implications [4, 12, 14–16]

S1	Increase recycling in reuse in VF
S2	Promote renewable energy sources for VF
S3	Increase efficient use of resources
S4	A partnership between state and civil society
S5	Use of political institutions to implement systematic change
S6	Retrofitting vacant urban buildings with green and smart building approaches for VF
S7	Focusing on promoting new job opportunities for VF
S8	Use of organic fertilizers
S9	End-to-end monitoring and control of VF systems with digital technologies
S10	Investing in growing anything additive to herbs and leafy greens
S11	Focusing on Niche production areas with VF

2.4 Methodological Background

This section will provide the methodological background with preliminary information about the 2-Tuple linguistic model and the HoQ of QFD. First, the main benefits and properties of the 2-Tuple model and Linguistic Hierarchies are given. Later, the primary steps of 2-Tuple integrated HoQ are provided with the general flow of the framework.

2.5 2-Tuple Linguistic Model

Herrera and Martinez first represent this model in 2000 [18]. The 2-tuple linguistic model and its extensions have been applied to various topics, mainly decision-making and decision analysis problems [18–20]. Basic definitions are as follows [18]:

The 2-Tuple fuzzy linguistic representation model represents the linguistic information using a 2-Tuple (S, α) here; S is a linguistic label, and α is a numerical value representing the value of the symbolic translation. The function is defined as:

$$\Delta_s : [0, g] \rightarrow \bar{S}$$
$$\Delta_s(\beta) = (S_i, \alpha),\ with \begin{cases} i = round(\beta) \\ \alpha = \beta - i \end{cases} \quad (1)$$

The linguistic term set S could be converted into 2-Tuple form by adding zero value as in the following relation:

$$S_i \in S \Rightarrow (S_i, 0) \quad (2)$$

For further details, the readers can refer to [18]. The main benefits of the 2-tuple linguistic model are the interpretability of the results, the possibility to deal with variables closer to the human beings' cognitive processes, and increased accuracy of computations. Regarding these benefits, to create a flexible environment for the DMs and a better analysis of the VF area, the suggested 2-Tuple methodology is integrated with the HoQ of QFD. The *Linguistic Hierarchies* [18] approach is used to unify the multigranular linguistic input under the one unified linguistic set. A transformation equation exists to normalize label sets with different granularity. The following equation gives the relations:

$$TF_{t'}^{t}\left(S_i^{n(t)}, \alpha^{n(t)}\right) = \Delta\left(\frac{\Delta^{-1}\left(S_i^{n(t)}, \alpha^{n(t)}\right) \cdot \left(n\left(t'\right) - 1\right)}{n(t) - 1}\right) \tag{3}$$

where *TF* is the transformation function for *Linguistic Hierarchies,* and the transformation is from tth level to t'th level.

2.6 2-Tuple Integrated HoQ

The 2-tuple integrated HoQ is suggested for this solution ranking problem. The QFD is initially introduced by Akao [21] in the late 60's, and it was introduced as a potent design tool for products. Its ability to reflect the end-user/customer needs into the design phase is the main advantage of the QFD method. Therefore, HoQ is the primary part of the QFD method, and this first matrix forms the basis of our suggested methodology. The details of the recommended methodology are provided in Fig. 1.

(1) Detecting VF challenges as CNs in HoQ framework, (2) Assigning weights of CNs according to DM preferences, (3) Detecting solutions to overcome CNs as DRs in HoQ framework, (4) Applying 2-Tuple integrated QFD framework [22] to obtain the solution prioritization for future of VF, are the main steps of the HoQ framework. In this paper, the HoQ method has been chosen thanks to its comfortable and robust computational steps, which can efficiently reflect the CNs' importance in DRs. Also, as mentioned in the 2-tuple QFD integrated studies, the technique is suitable to handle design and MCDM problems. Plus, by integrating HoQ with 2-tuple, this study has provided a flexible decision-making environment to DMs about the forms and interpretability of their judgments.

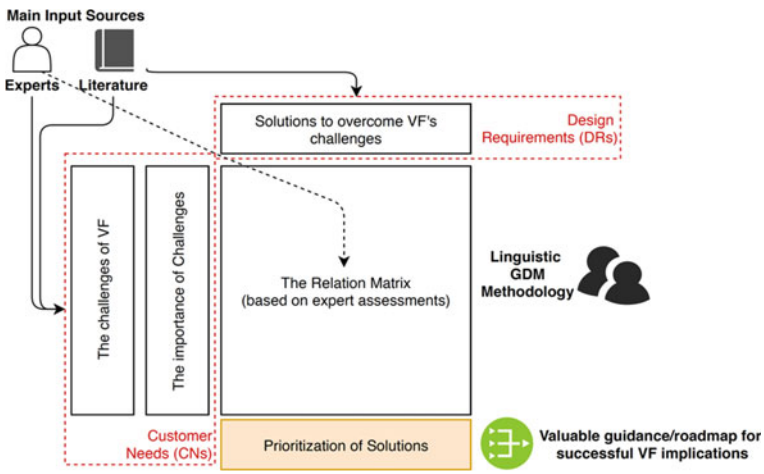


Fig. 1 Suggested framework

3 Case Study

The case study is designed to test the applicability of the suggested framework. When the framework is formed, the first step was creating a decision-making group and validating the framework to them.

The decision-making group is composed of three experts who are competent in smart agriculture, smart system design, and smart supply chains and their design. Due to the volatile background of DMs, two different linguistic sets are provided. The DMs works on smart agriculture and smart system design are assessed by using nine scaled linguistic sets. The third DM works on smart supply chains and their design assessed with five scaled linguistic sets. The details of linguistic sets are as follows:

Second level five scaled (Very Low (VL)—Low (L)-Medium (M)-High (H)-Perfect (P)) and the third level nine scaled (Very low (VL)—Low (L)-Medium Low (ML)-Almost Medium (AM)-Medium (M)-Medium High (MH)-High (H)-Very High (VH)-Perfect (P)) hierarchy of letters [18].

The Delphi [23] method is followed for the separate meetings with each DM. Their judgments are collected separately and aggregated by using the 2-tuple model's Linguistic Hierarchies approach. The details of the calculations are presented followingly:

1. As a first step, the challenges are gathered, as mentioned in Sect. 2.3.
2. Each DM assesses the importance of challenges, and their evaluations are unified and aggregated.
3. The solutions (Sect. 2.3) are derived from the literature and validated by experts.

4. As the third step, the assessments of each DM are unified and aggregated to apply 2-tuple HoQ to obtain solutions' priorities. Here Table 3 presents the aggregated decision matrix with aggregated importance of challenges.

4 Results and Analysis

Table 3 gives the aggregated relation matrix for the VF solution prioritization problem, and the weights of solutions are provided in the last row of the table. Figure 2 gives the results and also the sensitivity analysis conducted to test the reliability and the replicability of the same problem under different CN weightings.

The sensitivity is performed to see how the ranking is changed when the importance of CNs change according to the main dimensions of challenges. Due to the page limitation, the details of the analysis will be given during the presentation.

5 Conclusions

To conclude, this paper presents a linguistic MCDM model to overcome the challenges for effective VF implications. In order to achieve this goal, the 2-tuple linguistic integrated HoQ method is suggested. The challenges are grouped under four different dimensions such as economic, political, environmental, and social. A sensitivity analysis is conducted to better analyze the effects of dimensions on the solution ranking. Solutions that cover most of the challenges are detected as the most appropriate ones. For further researches, the number of solutions and the number of DMs can be augmented. Also, same model can be extended to be applied in other sectors such as health and urban design.

Table 3 Aggregated HoQ matrix

	Imp. of CNs	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
C1	(VH, 0.17)				(L, 0.22)	(L, 0.22)			(L, 0.22)		(P, 0)	(AM, −0.22)
C2	(H, −0.39)	(ML, 0)	(ML, 0)	(AH, −0.22)	(L, 0.22)		(L, 0.22)		(L, 0.22)	(AH, − 0.22)		
C3	(P, −0.43)	(AH, −0.22)		(L, 0)	(M, 0)		(P, 0)			(L, 0.22)		
C4	(P, 0)	(VH, −0.22)	(P, 0)		(L, 0.22)		(H, 0)					
C5	(AM, 0.17)		(L, 0.22)				(VH, 0.17)		(AH, −0.22)	(AH, 0.17)	(H, 0)	(P, 0)
C6	(H, −0.04)				(H, 0)				(M, 0)			(P, 0)
C7	(L, 0.22)				(L, 0.22)	(M, 0)	(P, −0.39)					
C8	(P, −0.39)				6.78	(H, 0)	(P, 0)	(AH, 0.17)				(ML, 0)
C9	(ML, 0.04)				(M, 0)	(AH, −0.22)		(VH, −0.22)				(ML, 0)
	Imp	(ML, 0.10)	(ML, −0.36)	(L, −0.29)	(AM, 0.16)	(L, 0.43)	(M, 0.31)	(L, 0.10)	(L, 0.38)	(L, 0.08)	(ML, −0.42)	(ML, 0.32)

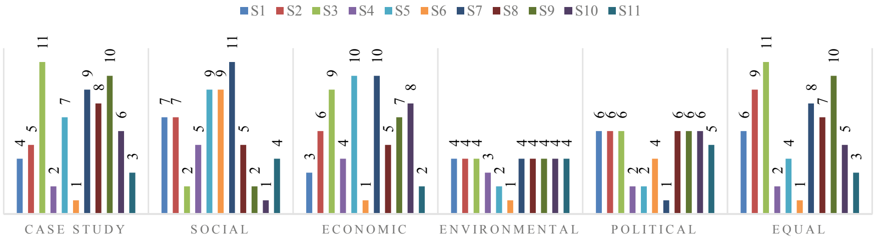


Fig. 2 Sensitivity analysis

Acknowledgements This work has been supported by the Scientific Research Projects Commission of Galatasaray University (Project Number: 19.102.002).

References

1. Graamans, L., Tenpierik, M., van den Dobbelsteen, A., Stanghellini, C.: Plant factories: reducing energy demand at high internal heat loads through façade design. Appl. Energy **262**, 114544 (2020)
2. Bai, C., Dallasega, P., Orzes, G., Sarkis, J.: Industry 4.0 technologies assessment: a sustainability perspective. Int. J. Prod. Econ. **229**, 107776 (2020)
3. Eden Green Technology: Home. https://ceac.arizona.edu/home (2021)
4. Benke, K., Tomkins, B.: Future food-production systems: vertical farming and controlled-environment agriculture. Sustain. Sci. Pract. Policy **13**, 13–26 (2017)
5. Despommier, D.: The vertical farm: controlled environment agriculture carried out in tall buildings would create greater food safety and security for large urban populations. J. Verbrauch. Lebensm. **6**, 233–236 (2011)
6. Cicekli, M., Barlas, N.T.: Transformation of today greenhouses into high-technology vertical farming systems for metropolitan regions. J. Environ. Prot. Ecol. **15**, 1066–1073 (2014)
7. Despommier, D.: Farming up the city: the rise of urban vertical farms. Trends Biotechnol. **31**, 388–389 (2013)
8. Molden, D.: International water management institute, comprehensive assessment of water management in agriculture (program). In: Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture. Sterling, VA, Earthscan, London (2007)
9. Kumar, A.R.: Precision agriculture: a review on its techniques and technologies **02**, 7 (2020)
10. Goodman, W., Minner, J.: Will the urban agricultural revolution be vertical and soilless? A case study of controlled environment agriculture in New York City. Land Use Policy **83**, 160–173 (2019). https://doi.org/10.1016/j.landusepol.2018.12.038
11. Al-Chalabi, M.: Vertical farming: skyscraper sustainability? Sustain. Cities Soc. **18**, 74–77 (2015)
12. Al-Kodmany, K.: The vertical farm: a review of developments and implications for the vertical city. Buildings **8**, 24 (2018)
13. van Asselt, S.: A Swot Analysis of the Vertical Farm (2018)
14. Briceno, D.G.-C.: Vertical Farming Sustainability and Urban Implications (2018)
15. Kotrokois, P.: Transforming Agriculture through Digital Technologies 14
16. Growing Beyond the Hype: Controlled environment agriculture. https://www.s2gventures.com/reports/growing-beyond-the-hype%3A--controlled-environment-agriculture
17. Anonymous: Vertical farm architectural challenge for self-sufficient urbanism. Space 98–99 (2008)

18. Martínez, L., Rodriguez, R.M., Herrera, F.: The 2-Tuple Linguistic Model. Springer International Publishing, Cham (2015)
19. Mi, C., Chen, Y., Zhou, Z., Lin, C.-T.: Product redesign evaluation: an improved quality function deployment model based on failure modes and effects analysis and 2-tuple linguistic. Adv. Mech. Eng. **10**, 1687814018811227 (2018)
20. Singh, A., Gupta, A.: Best criteria selection based PROMETHEE II to aid decision-making under 2-tuple linguistic framework: case-study of the most energy efficient region worldwide. Int. J. Manag. Decis. Mak. **19**, 44–65 (2020)
21. Akao, Y.: QFD: Integrating Customer Requirements into Product Design. Cambridge, MA (1990)
22. Mei, Y., Liang, Y., Tu, Y.: A multi-granularity 2-tuple QFD method and application to emergency routes evaluation. Symmetry-Basel **10**, 484 (2018)
23. Buyukozkan, G., Karabulut, Y., Arsenyan, J.: RFID service provider selection: an integrated fuzzy MCDM approach. Measurement **112**, 88–98 (2017)

New Product Development and Innovation Management

A Study of Perceived Level of Difficulty for the Execution of the New Product Development Process

Kwai Hong Lui Lucas

Abstract The current study would like to assess the perception of the staff involved in new product development, while the research investigation was conducted by a questionnaire-based survey configuration with the corresponding data analysis. Practically, the study provides the assessment about the perceived difficulties of the stages of the new product development process, namely, idea generation, idea screening, concept development and piloting, development of marketing strategy, business analysis, product development, testing the market, and commercialization. Along with the general demographic information of the engineers, including age, gender, level of education, rank, and years of service, the research questionnaire was composed to collect the required data for the exploration of the relationship between them can be revealed. With the self-defined research instrument, 132 valid responses were collected for the modelling process. After the completion of the analysis, testing the market was found to be the most difficult, and the particular difficulties for concept development and piloting and business analysis can be observed from the junior staff. With respect to the situation, the recommendations were provided in the latter part of the paper.

Keywords New product development · Perceived difficulty · Perception

1 Introduction

This study assesses the perceived difficulties of the stages of the new product development process so that the wise allocation of training resource can be accomplished. The new product development process is one of the most important processes for the manufacturing firms to maintain competitive advantage and have sustainable development along with the smooth application of innovation. Along with the empirical experience to the research and development process, it can be realized that, apart from the infrastructure and the hardware support of the research and development,

K. H. L. Lucas (✉)
Industrial & System Engineering, The Hong Kong Polytechnic University, Hong Kong, China
e-mail: lucas.lui@connect.polyu.hk

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_64

the perception of the difficulty to the process in the mind of the engineers would be deterministic to the overall performance of the process, while it may overwhelm the efficiency of the sophisticated equipment support [1]. Therefore, the current study would like to assess the perception of the staff involved in new product development. As the perceived difficulties for rolling out the new product development process hesitated the staff to contribute with their innovation, the wise allocation of training resource to the hardest stage can help to smoothen the process. Consequently, the comprehension of the through of the staff about the levels of difficulty for the stage was a must for the provision of proper training, while the current study would be dedicated to the associated investigation.

Along with the motivation, the aim of the current study was defined as to provide an assessment of the perceived level of difficulty for the execution of the new product development process so that practical recommendations for training can be generated. With the intention of completing the aim, the four objectives were defined: To review the stages of the new product development process; To order to the level of difficulty for each stage of the process; To conduct a segmented analysis in order to reveal the diversity of the opinion among different demographic groups; To give recommendations based on the analytic findings.

2 Methodological Approach

Along with the typical configuration of a questionnaire-based study, five steps were included. (1) The research instrument should be developed to collect information from the respondents. (2) The consent should be gained from the potential respondents before starting the questionnaire-based survey. (3) The questionnaire should distribute and gather from the respondents. (4) The data analysis would be conducted after obtaining the data. (5) The recommendations can be generated to direct the proper allocation of the research in dealing with the situation. The designed questionnaire [2] should consist of two parts so as to align with the needs. On the one hand, the demographic information, including age, gender, level of education, rank, and years of serving, would be assessed by the multiple-choice questions. On the other hand, the level of difficulties for the eight stages would be measured by the positively coded statements, as shown in Table 1. In essence, the respondents would be asked to show their level of agreeing to the corresponding statement via the 7-point Likert scale [3], while the selected option expressed the intensity of the related difficulties. Additionally, the related options would be converted to the numerical scale in order to enable statistical analysis in the following research process.

Two methods, namely, descriptive analysis [4] and the analysis of variance (ANOVA) [5], would be applied in the current study. Practically, the descriptive analysis would be used to reveal the demographic distribution of the respondents and the visualisation of the perceived difficulties, whereas the ANOVA was used to reveal the diversity of the opinion among different demographic groups.

Table 1 Statements used in assessing the difficulties of the stage of the new product development process

Stages	Positively Coded Statement
S1 Idea generation	1. From my experience of working for the new product development process, I think the idea generation procedure is very difficult in the development of a new product
S2 Idea screening	2. From my experience of working for the new product development process, I think the idea screening procedure is very difficult in the development of a new product
S3 Concept development and piloting	3. From my experience of working for the new product development process, I think the concept development and piloting procedure is very difficult in the development of a new product
S4 Development of marketing strategy	4. From my experience of working for the new product development process, I think the development of a marketing strategy is very difficult in the development of a new product
S5 Business analysis	5. From my experience of working for the new product development process, I think the business analysis procedure is very difficult in the development of a new product
S6 Product development	6. From my experience of working for the new product development process, I think the product development procedure is very difficult in the development of a new product
S7 Testing the market	7. From my experience of working for the new product development process, I think testing the market is very difficult in the development of a new product
S8 Commercialisation	8. From my experience of working for the new product development process, I think the commercialisation procedure is very difficult in the development of a new product

3 Findings and Discussion

Subsequent to the development of the methodological approach of the study, the questionnaire-based study was conducted, while 132 out of 150 valid responses were collected. Along with the checking of the reliability with Cronbach's alpha, while the reliability of the data was supported (Alpha Value = 0.884 > 0.7) [6, 7]. Additionally, the results for the KMO and Bartlett's Test were provided. In essence, the associated findings shown the significance of the collected data (Sig. = 0.001 < 0.05), while the corresponding results can be seen in Table 2.

Table 2 Summarized results of KMO and Bartlett's test for the current project. Lucas: keep Table 4

KMO and Bartlett's test		
Kaiser–Meyer–Olkin measure of sampling adequacy		0.903
Bartlett's test of sphericity	Approx. Chi-Square	440.155
	df	28
	Sig.	0.000

Table 3 Cross-correlation of the level of difficulties for the eight stages of the new products development process

	S1	S2	S3	S4	S5	S6	S7	S8
S1	1	0.490**	0.551**	0.480**	0.553**	0.435**	0.490**	0.557**
S2	0.490**	1	0.494**	0.534**	0.449**	0.471**	0.439**	0.346**
S3	0.551**	0.494**	1	0.525**	0.452**	0.497**	0.486**	0.467**
S4	0.480**	0.534**	0.525**	1	0.471**	0.487**	0.510**	0.435**
S5	0.553**	0.449**	0.452**	0.471**	1	0.529**	0.406**	0.531**
S6	0.435**	0.471**	0.497**	0.487**	0.529**	1	0.578**	0.529**
S7	0.490**	0.439**	0.486**	0.510**	0.406**	0.578**	1	0.503**
S8	0.557**	0.346**	0.467**	0.435**	0.531**	0.529**	0.503**	1

** Correlation is significant at the 0.01 level (2-tailed)

3.1 Descriptive Analysis

Next, the average scores of the perceived difficulty for each stage would be presented, while the descending order of the perceived difficulties can be produced, while the sequence was: Testing the Market (Mean Score = 4.05) > Idea Screening (Mean Score = 4.01) > Commercialization (Mean Score = 3.95) > Concept Development and Piloting (Mean Score = 3.87) > Business Analysis (Mean Score = 3.85) > Product Development (Mean Score = 3.83) > Idea Generation (Mean Score = 3.78) > Development of Marketing Strategy (Mean Score = 3.66). Therefore, the allocation of the resource in remediating the situation can be wised allocated. In addition, the cross-correlation of the eight stages was provided, while the related results can be seen in Table 3.

3.2 Analysis of Variance

Finally, the analysis of variance (ANOVA) for the perceived difficulties with respect to the demographic information would be conducted in order to reveal the diversity of the opinion among the respondents. First, the analysis with respect to age would be conducted, while the associated results can be seen in Table 4. From

Table 4 Result of the ANOVA of eight stages of the new product development process with respect to age

		Sum of squares	Degree of freedom	Mean square	F	Significant value
Idea generation	Among groups	43.430	7	6.204	1.342	0.236
	Inside groups	573.199	124	4.623		
	Total	616.629	131			
Idea screening	Among groups	42.536	7	6.077	1.312	0.250
	Inside groups	574.456	124	4.633		
	Total	616.992	131			
Concept development and piloting	Among groups	87.165	7	12.452	2.926	0.007
	Inside groups	527.646	124	4.255		
	Total	614.811	131			
Development of marketing strategy	Among groups	49.450	7	7.064	1.575	0.149
	Inside groups	556.209	124	4.486		
	Total	605.659	131			
Business analysis	Among groups	88.141	7	12.592	2.804	0.010
	Inside groups	556.829	124	4.491		
	Total	644.970	131			
Product development	Among groups	48.322	7	6.903	1.416	0.205
	Inside groups	604.671	124	4.876		
	Total	652.992	131			
Testing the market	Among groups	58.170	7	8.310	1.828	0.087
	Inside groups	563.557	124	4.545		
	Total	621.727	131			
Commercialisation	Among groups	18.563	7	2.652	0.543	0.801
	Inside groups	606.065	124	4.888		
	Total	624.629	131			

the numerical result, it can be seen that the diversity can be observed for Concept Development and Piloting (Sig. $= 0.07 < 0.05$) and Business Analysis (Sig. $= 0.010 < 0.05$), while the shared opinion can be observed from the rest of the perceived difficulties (Sig. = Range from 0.087 to $0.801 > 0.05$).

As diversity can be seen from Concept Development and Piloting and Business Analysis, the mean analysis would be conducted, while the corresponding result can be observed from Table 5. It can be seen that the young staff (Less than 25 Years Old) considered the two stages were the most difficult part of the new product development process.

Additionally, the results of Scheffe analysis of concept development and piloting and business analysis would be provided in Table 6, respectively.

Additionally, no observable diversity of the opinion can be seen in order demographic segmentation.

4 Conclusion and Recommendations

On the one side, two findings can be concluded, while they are: (1) The order of the stage with respect to the level of difficulty was: Testing the Market > Idea Screening > Commercialisation > Concept Development and Piloting > Business Analysis > Product Development > Idea Generation > Development of Marketing Strategy. (2) The young staff has particular difficulties in concept development and piloting and business analysis. On the flip side, two recommendations would be proposed to remediate the situation. (1) The enforced training can be provided to the staff about

Table 5 Result of mean analysis of eight stages of the new product development process with respect to concept development and piloting and business analysis

Age		Concept development and piloting	Business analysis
Less than 25 years old	Mean	5.10	5.35
	N	20	20
	Std. Deviation	1.774	2.059
26–30 years old	Mean	3.72	3.11
	N	18	18
	Std. Deviation	2.270	2.423
31–35 years old	Mean	3.16	3.42
	N	19	19
	Std. Deviation	2.218	1.805
36–40 years old	Mean	3.42	4.00
	N	19	19
	Std. Deviation	2.457	2.357
41–45 years old	Mean	5.00	4.56
	N	9	9
	Std. Deviation	2.121	2.128
46–50 years old	Mean	4.24	3.88
	N	17	17
	Std. Deviation	1.954	1.996
51–55 years old	Mean	4.06	2.65
	N	17	17
	Std. Deviation	1.919	1.902
Over 55 years old	Mean	2.38	4.00
	N	13	13
	Std. Deviation	1.502	2.236
Total	Mean	3.87	3.85
	N	132	132
	Std. Deviation	2.166	2.219

Table 6 Result of Scheffe analysis with respect to piloting and business analysis

(I) Age	(J) Age	Mean difference (I-J)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Less than 25 years old	26–30 years old	2.239	0.688	0.170	−0.39	4.87
	31–35 years old	1.929	0.679	0.335	−0.66	4.52
	36–40 years old	1.350	0.679	0.783	−1.24	3.94
	41–45 years old	0.794	0.851	0.996	−2.45	4.04
	46–50 years old	1.468	0.699	0.731	−1.20	4.14
	51–55 years old	2.703*	0.699	0.045	0.03	5.37
	Over 55 years old	1.350	0.755	0.864	−1.53	4.23
26–30 years old	Less than 25 years old	−2.239	0.688	0.170	−4.87	0.39
	31–35 years old	−0.310	0.697	1.000	−2.97	2.35
	36–40 years old	−0.889	0.697	0.977	−3.55	1.77
	41–45 years old	−1.444	0.865	0.902	−4.75	1.86
	46–50 years old	−0.771	0.717	0.991	−3.51	1.97
	51–55 years old	0.464	0.717	1.000	−2.27	3.20
	Over 55 years old	−0.889	0.771	0.987	−3.83	2.06
31–35 years old	Less than 25 years old	−1.929	0.679	0.335	−4.52	0.66
	26–30 years old	0.310	0.697	1.000	−2.35	2.97
	36–40 years old	−0.579	0.688	0.998	−3.21	2.05
	41–45 years old	−1.135	0.857	0.971	−4.41	2.14

(continued)

Table 6 (continued)

(I) Age	(J) Age	Mean difference (I-J)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
	46–50 years old	−0.461	0.707	1.000	−3.16	2.24
	51–55 years old	0.774	0.707	0.991	−1.93	3.48
	Over 55 years old	−0.579	0.763	0.999	−3.49	2.33
36–40 years old	Less than 25 years old	−1.350	0.679	0.783	−3.94	1.24
	26–30 years old	0.889	0.697	0.977	−1.77	3.55
	31–35 years old	0.579	0.688	0.998	−2.05	3.21
	41–45 years old	−0.556	0.857	1.000	−3.83	2.72
	46–50 years old	0.118	0.707	1.000	−2.58	2.82
	51–55 years old	1.353	0.707	0.816	−1.35	4.06
	Over 55 years old	0.000	0.763	1.000	−2.91	2.91
41–45 years old	Less than 25 years old	−0.794	0.851	0.996	−4.04	2.45
	26–30 years old	1.444	0.865	0.902	−1.86	4.75
	31–35 years old	1.135	0.857	0.971	−2.14	4.41
	36–40 years old	0.556	0.857	1.000	−2.72	3.83
	46–50 years old	0.673	0.874	0.999	−2.66	4.01
	51–55 years old	1.908	0.874	0.687	−1.43	5.25
	Over 55 years old	0.556	0.919	1.000	−2.95	4.07
46–50 years old	Less than 25 years old	−1.468	0.699	0.731	−4.14	1.20
	26–30 years old	0.771	0.717	0.991	−1.97	3.51
	31–35 years old	0.461	0.707	1.000	−2.24	3.16

(continued)

Table 6 (continued)

(I) Age	(J) Age	Mean difference (I-J)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
	36–40 years old	−0.118	0.707	1.000	−2.82	2.58
	41–45 years old	−0.673	0.874	0.999	−4.01	2.66
	51–55 years old	1.235	0.727	0.893	−1.54	4.01
	Over 55 years old	−0.118	0.781	1.000	−3.10	2.86
51–55 years old	Less than 25 years old	−2.703*	0.699	0.045	−5.37	−0.03
	26–30 years old	−0.464	0.717	1.000	−3.20	2.27
	31–35 years old	−0.774	0.707	0.991	−3.48	1.93
	36–40 years old	−1.353	0.707	0.816	−4.06	1.35
	41–45 years old	−1.908	0.874	0.687	−5.25	1.43
	46–50 years old	−1.235	0.727	0.893	−4.01	1.54
	Over 55 years old	−1.353	0.781	0.883	−4.34	1.63
Over 55 years old	Less than 25 years old	−1.350	0.755	0.864	−4.23	1.53
	26–30 years old	0.889	0.771	0.987	−2.06	3.83
	31–35 years old	0.579	0.763	0.999	−2.33	3.49
	36–40 years old	0.000	0.763	1.000	−2.91	2.91
	41–45 years old	−0.556	0.919	1.000	−4.07	2.95
	46–50 years old	0.118	0.781	1.000	−2.86	3.10
	51–55 years old	1.353	0.781	0.883	−1.63	4.34

* Correlation is significant at the 0.05 level (2-tailed)

the support of testing the market. For instance, training for the use of historical data for decision making can be provided [8]. (2) For the junior staff, the training for the automated process to simulate and to visualise the concepts [9] can be provided, while the simulation package for the business operation can be included in the curriculum of the staff orientation [10].

References

1. Song, X.M., Parry, M.E.: A cross-national comparative study of new product development processes: Japan and the United States. J. Mark. **61**(2), 1–18 (2018)
2. Hill, H., Litke, E., Lynch, K.: Learning lessons from instruction: descriptive results from an observational study of urban elementary classrooms. Teach. Coll. Rec. **120**(12) (2018)
3. Hamed, A.M., Mozhgan, D., Reyhaneh, S., Reyhaneh, S.: Designing, constructing and analyzing likert scale data. J. Educ. Commun. Health **5**(3) (2018)
4. Melkman, E., Hershkowitz, I., Zur, R.: Credibility assessment in child sexual abuse investigations: a descriptive analysis. Child Abuse Negl. **67** (2017)
5. Pandis, N.: Analysis of variance to linear regression. Am. J. Orthod. Dentofac. Orthop. Official Publ. Am. Assoc. Orthodontists IT Constituent Soc. Am. Board Orthod. **149**(6) (2016)
6. Dominguez-Lara, S.A., Merino-Soto, C.: A modification of Cronbach's alpha coefficient by correlated errors. Revista Medica De Chile **145**(2) (2017)
7. Statistics How To: Cronbach's alpha: simple definition, use and interpretation. Retrieved 20 Sept 2020, from https://www.statisticshowto.com/cronbachs-alpha-spss/ (2020)
8. Mitroff, I.I.: Technology Run Amok: Crisis Management in the Digital Age. Springer International Publishing AG, Cham (2018)
9. Upadhyay, S., McCormick, K.: The Revenue Acceleration Rules: Supercharge Sales and Marketing Through Artificial Intelligence, Predictive Technologies, and Account-Based Strategies. John Wiley & Sons, Inc. (2018)
10. Fan, G.-F., Li, M.-W., Hong, W.-C.: Short-term load forecasting by artificial intelligent technologies. In: Short-Term Load Forecasting by Artificial Intelligent Technologies. MDPI (2019)

Is that Innovation? Unlocking Vietnamese MedTech SMEs Innovation Pathways

Leanne Chung, Kim Hua Tan, and Thi Thu Thuy Nguyen

Abstract In the 24th ICPR, we unraveled the unique Chinese innovation pathways and clearly explained its distinctiveness from the Japanese and South Korean. In this paper, we examine SMEs innovation pathways using empirical data gathered in Vietnam. Under current America's 'decoupling' strategy, Vietnam will benefit greatly from manufacturing firms moving out of China. Vietnam is now a major destination of foreign direct investment in research and development and an attractive knowledge-based location for leading MNCs. Many MNCs (such as GE Healthcare, Philips, Siemen, Hitachi, Mindray) have set up their R&D units in Hanoi and/or Ho-Chi-Minh. Hence, a better understanding of Vietnamese innovation in the context of Vietnam's emerging economy, evolving institution, and growing firm capabilities is beneficial to practitioners, policy makers and academia. We explore the unique innovation pathways from the perspective of Vietnamese medical equipment manufacturers (MedTech) and provide insight from managers on how international firms could galvanized the uniqueness to their product development advantage. The results provided interesting insights into how MedTech (a high growth sector in Vietnam) SMEs cultivated their data analytics, collective and target innovation, digitalization skills and know how in various innovation phases. Vietnam is not a paragon of innovation at all, but it certainly has learned how to invent sustainably with less, and that is something other emerging and Asian economies can learn from. We believe the findings from our empirical case studies can contribute towards a better understanding of how Vietnam has been able to evolve from a position of technology borrower to technology innovator. Moreover, findings from this research help to shed light on existing Asian innovation development debates.

L. Chung (✉)
Cardiff Business School, Cardiff, England
e-mail: Chungl1@cf.ac.uk

K. H. Tan
Nottingham University Business School, Nottingham, England
e-mail: Kim.tan@nottingham.ac.uk

T. T. T. Nguyen
MISA JSC, Hanoi, Vietnam

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_65

Keywords Innovation management · Vietnam · Product development · Creativity · Healthcare

1 Introduction

MedTech innovation in Vietnam has been progressing steadying amid fragmented, unorganized and mostly done in conjunction with government support. However, lately a steady stream of Vietnamese MedTech products is making entry into the international market. In 2020, Vietnam made global headlines for its effective management of COVID-19 [1]. Effective control such as extensive contact tracing, isolation and targeted lockdown is vital but one of the key factors determining the success of Vietnam's approach has been the ability of the country MedTech SMEs to rapidly increase its diagnostic capacity through rapid innovation of mass screening products. For example, upon the first case in January 2020, within two months, first locally produced COVID-19 test kits were developed and commercialised by a MedTech SME. The kit has now been widely used with WHO recommendation [1, 2].

While the ultimate outcome from the COVID-19 disruptions remains to be seen, we believe valuable lessons could be learnt by studying and analysing the rapid success of Vietnam's MedTech SME innovation pathways. The timely development of effective and affordable virus test kits in Vietnam offers especially important lessons, especially with respect to how emerging economies can foster health-focused innovation rather than relying on the diffusion of innovation from abroad. Vietnam is a lower to middle-income country with a population of 97 million. In 2018, its gross domestic product per capita was US$2.566, and around US$150 per person were spent on healthcare [1]. The healthcare system in Vietnam includes both public and private providers, with less than one doctor and two nurses per 1000 population. Public hospitals are divided into four levels: central, provincial, district and commune levels, and are responsible for providing healthcare for the majority of the residents in Vietnam [3]. Vietnam central government has set to achieve upper-middle-income country by 2025.

The ongoing US and China trade war also helps Vietnam to notch in a handful of high-profile investments. In 2020, Vietnam attracted Pegatron, a supplier for Apple and Samsung, to LG Electronics, which relies on Vietnam for its vehicle and smartphone businesses [4]. By the end of the year, nearly all of Apple's major suppliers in the region had set up shop in Vietnam or planned to do so. The relocations continue a "China Plus One" trend among companies that have reduced their reliance on Asia's biggest economy because of rising costs, the risks from a trade war with the US, and Covid-19 disruptions to the supply chain [4]. Hence, Vietnam is also increasingly recognized as a major source of product technology innovation as well as for its innovative business models and unique healthcare ecosystem. Nevertheless, success in Vietnam does not come easy, it requires agility, innovation, and long-term commitment to stay abreast of trends in this fiercely competitive market. To take the pulse of Vietnam unique innovation pathway, and to understand its challenges, opportunities, and priorities, case study with 8 MedTech SMEs was conducted and selected executive interviews. This paper summarizes insights and findings on how MedTech SMEs can thrive amid the challenging and complex external environment in Vietnam, step up on innovation, commercial, and strategic capabilities, and embrace the next wave of growth in the region.

2 SMEs Challenges and Innovation Pathways in Asia

Innovation is widely accepted as the important factor to create competitiveness of both firms and nations [5–8]. The role of innovation is more important in the context of SMEs, and significantly increasing in the current situation of globalization trend, industry 4.0, shortened product life cycles and rapid changes in demand.

Most of the Asian SMEs built their own products for a niche market. Hence, innovation is seen as the core business strategy for SMEs to increase their survival rate, development potential and productivity [9]. If SMEs do not consider innovation as the core business strategy, they may not survive or become uncompetitive. The reasons may come from the obsolescence of their products or services. Therefore, Chung and Tan [9] point out that SMEs that do innovation can outperform those that do not. Due to the small firm size and the ability of fast response to market trends, SMEs are capable to implement more efficient new product development than large firms [10]. They can connect with MNCs or conduct outsourcing and acquire technology and knowledge along the process. Researchers described these processes as adaptation, imitation or technology mastery to reduce cost and shorten time [11].

Japan, South Korea and China SMs provided many successful examples on how to turn imitation to innovation. The innovation pathways of for Japan, South Korea and China is shown in Table 1. In Japan, the three key stages concept of 'Shu-Ha-Ri' [12] has impacted many generations in Japan. In the initial stage, students should learn from the masters and faithfully obey them. Subsequently, once the students are able to master the skills proficiently, they should break away from the 'old' practices gradually. Finally, the students would depart from the tradition teachings to establish their own values. However, the learning process is more than just a simple imitation. During the imitation process, Japanese firms will select and master the useful skills independently, while improving and transforming the basic technologies into new products. The continuous improvement in products and business processes (i.e. *kaizen*) were achieved through 'grass-root' company-wide and 'collective' efforts.

In the early stage of South Korea industrialization, many firms (i.e. Samsung, Hyundai and LG) engaged in foreign joint ventures (JVs) for learning and securing access to new technologies. Kim [13] revealed that South Korea's industrialization had undergone three stages, which are duplicative imitation stage, creative imitation stage and innovation stage. In duplicative imitation stage, South Korean companies strived to obtain advanced technologies through formal mechanisms such as

Table 1 Comparison of Japan, South Korea and China innovation pathways

Innovation pathways	Descriptions	Refs.
Japan 守 (SYU)⇨破 (HA)⇨離 (RI)	Syu: Learn skills and knowledge from the masters, then obey them in a faithful way Ha: Break away from the old technologies, practices, skills and knowledge gained in Stage 1 and gradually invent their new products Ri: Leave the traditional technologies and knowledge behind and create new own values	[12]
South Korea Duplicative imitation ⇨ Creative imitation ⇨ Innovation stage	Duplicative imitation: Learn new skills and knowledge via two methods: formal one by technical license or informal by technical or literature assistance Creative imitation: Improve product values improve internal technology capabilities by cooperating with research organizations, local universities or cooperate with foreign companies Innovation: Heavy R&D investment, strengthen collaborations with research organizations, and recruit high-quality researchers globally	[13]
China 引 Yin ⇨ 调 Tiao ⇨ 创 Chuang	Yin: Implement imitation using local technology and resources, with limited resources and support from Western economies Tiao: Use the glocalisation strategy to add new features or reduce some functions to current products to test market Chuang: Transferred from local product to universal products which can be accepted in the global market	[9]

technical licensing agreements or informal mechanisms such as literature and technical assistance. In the creative imitation stage, South Korean firms attempted to reach higher level by gradually improving their products values. They do so by cooperating with local universities and research institutions. Moreover, they also fostering local technical talents, carrying out internal technology research and development, building strategic alliance with foreign companies. Eventually, the South Korean entrepreneurs made major shift from imitating advanced technologies to developing innovation technologies by themselves after years of establishing their knowledge-building mechanisms through employees training, foreign technology transfer and the diffusion of new knowledge among all the technical workers across the companies.

Chinese firms have more resources to support product innovation; i.e. larger pool of qualified personnel, sufficient raw materials, and good collaborative relationships with foreign organizations [9]. Hence, based on such a strong economic and political environment, Chinese firms may have their own unique product innovation trajectory compared to Japan and South Korea. The initial product innovation development stage for Chinese firms can be named as "引**(Yin)**". The term "Yin" in Chinese has the meaning of introducing and adopting others' knowledge or experience to achieve a better result. Likewise, the next product innovation development phase is "调**(Tiao)**". It means a way to adjust and improve the existing products. From the viewpoint of Chinese firms, this step is seen as a transition stage between imitation and innovation. According to Chung and Tan [9], Chinese firms usually like to test the market reaction by adding a number of new features on the existing products. Hence, Chinese firms at the Tiao phase is closer to innovating new products than the stages of "creative imitation" and "breakdown" in South Korea and Japan respectively. The final stage is innovation "创(**Chuang**)", which denotes the meaning of altering the original designs and creating new products for more profit.

Japan, South Korea and China have implemented effectively follower strategies to acquire and develop new technologies and knowledge to build up a knowledge economy. From that, they created the new product with their own brand to compete successfully in global market. Therefore, there should be many lessons for other developing countries in Asia like Vietnam to learn in order to develop their own innovation pathways. However, each country will have a specific situation and background which would have a considerable effect in the innovating process.

Overall, innovation in Vietnam has been progressing but still fragmented, unorganized and mostly done in conjunction with government organizations. However, some of Vietnamese brands were known in the international market. Although Vietnam can learn from Japan, South Korea and China, there are some differences in the innovation pathways. Vietnam, as a developing country, have been under resource-constrained and mainly developing low-cost products and solutions to meet local demands. In addition, the Vietnamese government has significant effects on firms' innovation. For the sake of comparison, Vietnam innovation pathways could be summed up in three stages, namely "Learning", "Developing" and "Innovating" [3].

The first stage is 'learning' it means that Vietnamese firms can absorb the existing technology and knowledge through license and training. This is similar to the situation of South Korea and Japan. This period started right after 'Doi Moi' in 1986. In this period, Vietnamese government concentrated on building to Vietnam's market as a cheap and young labor market in order to attract investment from foreigner corporations [3]. The second stage is the "developing". After 30 years of Doi Moi, Vietnam has achieved some positive results, which help Vietnam has become an emerging economy in area. In this period, together with the Industry 4.0 with the development of Artificial intelligence, Internet of Thing and Big Data, Vietnamese firms are rushing to catch up with the world trend. In this stage, especially in the industry of smart phone, Vietnam has their own brand of Bphone launching the first time in 2015 and the second time in 2017. The third stage is "innovating". In this stage, there may be the time Vietnamese firm can use their own technology to produce a new product. Typical examples are the MedTech SMEs that do research and development by themselves and introduce "made in Vietnam" products to local and international market. However, the characteristic of the unique Vietnamese pathways is not fully understood and need researching.

3 Research Methodology

The case study method will be adopted to study the MedTech SMEs innovation pathway. One of the main reasons is that MedTech innovation is a relatively new concept for Vietnamese SMEs. As a result, quantitative modeling or survey seems to be inappropriate. Qualitative method such as case study is more commonly used in exploratory studies [14]. Case study enables researchers to structure a new perspective through participants' knowledge and allows a deeper understanding of problems [15].

One of the authors is Vietnamese and has more than 6 years MedTech working experiences in Hanoi. Primary data was collected through semi-structured interviews that were conducted via Messenger and Zalo applications in Vietnamese. In order to protect the privacy of respondent's information, we code the respondents as R1–R8 (see Table 2). For each case, we selected CEO, business manager or product manager for the interview. This is because they have good knowledge of product development (ranges from 5 to 15 years) as well as overall business strategies and MedTech trends in Vietnam. Respondents R6, R7, R8 are from three big companies in Vietnam. The reason is that although they are currently working in large companies, but they have had many working experiences in small and medium companies for years. Therefore, their knowledge and insight into SME innovation are also valuable.

The interview questions were structured into three main topics. The first is the background of the correspondent and general information of the company. The second topic was on the interviewees' perspective on the unique characteristics of innovation pathway in MedTech. The third topic was on the innovation challenges and opportunities in MedTech. The average time duration of each interview was 40–90 min. To complement the interview, secondary data were collected from Vietnam's annual health reports, financial reports as well as product development plans of interviewed companies and government regulations on MedTech. All interviews were recorded and transcribed. Thematic analysis was carried out to identify the main themes, similarities and differences between companies, and key innovation ideas and concepts.

4 Findings and Discussion

Vietnam's medical equipment market has grown over the last 20 years with significant changes in its business and in the product development process. Following are some comments from respondents:

> *R2: In the early 20 years, Vietnam's healthcare market was essentially a "playground" for state-owned companies. The state buys medical equipment under the advice of experts in this industry. At the same time, these companies do not have many comparisons of quality, price, technical features and the modern level of equipment. In the context of a poor country and the internet was not available, the medical device market was mainly using aid from*

Table 2 Case companies information

Respondent	Position	Company background	SME[a] type
R1	CEO	– Main products: imaging diagnosis, biochemical – Revenue: VND 7 billion; number of staffs: 8	Super small
R2	Medical division head	– Main products: imaging diagnosis, intensive care, surgical devices – Revenue: VND 100 billion; Number of staffs: 30	Small
R3	Business manager	– Main products: intensive care, sterilizer – Revenue of equipment team: VND 100 billion; number of staffs: 10	Small
R4	Product manager	– Main products: imaging diagnosis – Revenue from medical device business: VND 250 million; number of staffs: 100	Medium
R5	Product and sale manager	– Main products: operating rooms, sterilization, endoscopy, ventilator, rehabilitation – Revenue: VND 300 billion; number of staffs: 30	Medium
R6	Product manager	– Main products: imaging diagnosis, endoscopy, surgery – Revenue: VND 700 billion; number of staffs: 60	Big
R7	Sales engineer	– Main products: hemodialysis, surgical equipment, 2 factories – Revenue: VND 1000 billion; Number of staffs: 1200	Big
R8	Product manager	– Main products: operating theaters, intensive care unit, sterilization – Revenue: VND 1000 billion; number of staffs: 100	Big

[a] *Note* The authors classify companies as per definition of Vietnamese article 6 of Decree 39/2018/ND-CP. According to this regulation, microenterprises have no more than 10 employees and total turnover of not more than 10 billion. Small businesses have a workforce from 11 to 50 people and revenues from 11 to 100 VND billion. Medium enterprises have 51–100 employees and turnover shall not exceed 300 billion

other countries. On the other hand, manufacturers do not know much about the Vietnamese market.

R3: In the past, when the market was high demand, low competition, most of the medical equipment business takes place between government and foreign partners. At that moment, most products were outdated technology than similar products in the world.

R2: The medical device market in Vietnam has been growing for 20 years. There are many companies dealing in medical devices in a variety of forms, such as state-owned companies, *private companies, FDI Company. However, most are small and medium-sized companies with a limited budget to invest in innovation, new product development.*

To better understand the product development process of MedTech SMEs, it is necessary to understand how their business model and customers. The key MedTech customers are hospitals or more specifically doctors and technicians. This group of clients can be divided into two main groups: public hospitals and private hospitals. In the case of public hospitals, they tend to use products originating in developed countries such as US, Japan, Germany. Especially the central hospital headed by country's leading specialist and doctors, most of the MedTech products must have G7 origin. For private hospitals, the origin of goods is more diverse. One of the managers pointed out that most of the MedTech SMEs are trading companies. They import equipment from overseas and then resell them to hospitals in Vietnam. The SMEs tend to focus on one or two specific MedTech products. The reasons come firstly from the nature of medical equipment which is complicated technology. Hence, training and knowledge transfer to master the machine applications and maintenance may take up to 3 months.

One of the managers pointed out that MedTech SMEs normally connected with big foreign firms to imitate their technologies. R7 commented *"Major medical equipment manufacturers in the world export goods to Vietnam. Then, they arrange specialists to transfer these technologies to SMEs in Vietnam"*. This collaboration can help SMEs reduce cost and shorten learning time. In this relationship, Vietnamese SMEs play the role of importers, distributors of foreign manufacturers, a member of a joint venture or cooperation to absorb technology. Sometimes, a firm maybe a distributor of two foreign competitors (for example, a distributor of both Mizuho and Getinge, both are operating table manufacturers). With this strategy, the SME will have to handle competition between two brands within a company. However, if the partnership is successful, it will help the company cover the entire market. One of the managers pointed out that this co-opetition can help SMEs to both create benefits and advance technological innovation i.e. develop cheaper version of operation tables for small private hospitals. *R3 pointed out that "Vietnamese SMEs only retrace the backward technology but it is still acceptable to gradually improve in the future."* However, these local innovated products are still very simple, low-end, and cheap imitation. They also set a quite low price in order to sell these niche products to district hospitals and private hospitals. These clients require a reasonable product quality with a competitive price. With these mindsets, some Vietnamese manufacturers have competed successfully.

Overall, there is optimism on MedTech innovation in Vietnam. Four key themes emerging from the case studies that characterize the unique innovation shift for MedTech SMEs (see Fig. 1). In the era of digital economy, data resources become an important strategic resource for SMEs. Massive data utilization has changed the way of production and global supply chain. Many leading MedTech MNCs (such as Hitachi, GE etc.) have gathered big data from their products and platforms to gain better insights of market needs and develop new products. Leveraging the value of big data will become the basis for competition for today's enterprises. Hence, the

	From	To	Characteristics
	Sharing customers' data to foreign MedTech providers	Not just sharing but also more effective data analytic to better understand local needs and demand	Gaining first mover advantage in developing niche products that meet customer needs
	Low innovation performance with the 'lone wolf' approach	Strengthened national and local collaboration mechanism for MedTech innovation	Effective collaboration is needed at all levels
	Waiting for new products from foreign firms to fill demands	Innovate low cost but effective products	Targeted innovation can meet local market demand
	Scramble for latest digital development and technologies	Finding inspiration in "digital" business models and new technologies from outside the healthcare arena	Leveraging digital technologies to revolutionize conventional market needs and behaviour.

Fig. 1 Four shifts in Vietnamese MedTech innovation

ability of Vietnamese MedTech SMEs to aggregate, elaborate and analyse the data is becoming a key competitive advantage. However, data sharing loses its unique right to data, and competitors' opportunism will cause great losses to the sharing party. One of the CEOs asked, "*How could SMEs share data with MNCs and maintain innovative to enhance their innovation performance under the condition of sharing data?*".

Most respondents pointed out that SME tends to be 'lone wolf' trying hard to understand and master a complex technology and ecosystem of MedTech. This is an uphill task both for product innovation as well as competing with leading high quality foreign brands. The high risk and high investment of innovation make SMEs shy away from product innovation, and the complexity and uncertainty of product innovation make them worry about their limited resources and capabilities. Hence, SMEs should not innovate in isolation, at least not in an effective way. To overcome the limitations, SMEs have to break through the organizational limitations and seek to cooperate with other in product innovation by building a unique SMEs community with passion for MedTech innovation and desire to help out. This effective collaboration push innovation beyond the reach of a lone wolf by tapping into collective insight and skill sets.

One of the managers pointed out that the old model: waiting for new products from foreign MNCs and develop cheap imitation is long gone. Those days are gone and the focus is rapidly shifting towards more effective targeting innovation to meet local needs, and gaining back control over customers from MNCs. *R7 commented that "Now, the biggest barrier for SMEs in Vietnam is not technology. It is easy to see that we are using modern medical equipment on a par with the world. Here are many inventions not only in the medical field but in other fields as well. However, these activities are fragmented and have not been in commercial. Owners of the products specialize only in technology, not the person with business thinking. To be able to turn those products into a commercial product, it requires a bridge between technology owners and financial institutions"*. Hence, beginning with a focus on the boundaries of existing core products, SMEs can improve their offerings through local innovation that can range from soft-touch development of existing products to innovative and ground-breaking new technologies that meet hospital needs.

All respondents also pointed out that SMEs are finding inspiration in "digital" business models and new technologies from outside the healthcare arena to provide a foundation for disruptive digital strategies that have the potential to revolutionize conventional market needs and behaviour. Digital business models should be adapted to fit the healthcare settings. One of the managers argued that MedTech SMEs should develop a framework for digital innovation in order to challenge themselves and their teams on how digital technologies can be used across their existing business operations, products and services. Innovative MedTech SMEs should draw parallels with leading technology firms in relation to their business models as well as adopting digital technologies for product innovation. SMEs are able to leverage digitization and stay relevant within the MedTech sector.

5 Conclusion

COVID-19 and its prolonged lockdowns have acted as an accelerant for a number of trends in healthcare, such as the dramatic shift to online consultation, evolving patient behaviors, and emergent of local MedTech innovation. Another big visible trend is the increased awareness of MedTech sourcing, supply chain and need of homegrown technologies to meet future disruptions. MedTech companies are now positioned to profit from significant opportunities in digitalization, particularly in Vietname where conditions are ripe for rapid digitalization to take place. The US-China trade-was also hasten the need for Vietnamese firms to step up innovation capabilities to take over China as the world's factory For emerging and advanced economies alike, the Vietnamese case serves as an important one to study, to understand ways in which affordable innovation can help thwart the devastation of pandemics.

Results from case studies show that in the face of these challenges, the classic imitates and innovate pathway of Japan, South Korea and China won't suffice in Vietnam. MedTech SMEs must embrace effective and targeted innovation to build a sustainable competitive advantage in the next normal. Playing catch-up can no longer be the main compass for change. Instead, transformations must take a truly holistic perspective (big data, collective, target innovation, and digital) to unlock new opportunities that SMEs have failed to harness at scale in the past, while enabling them to adapt at a much higher speed.

To be successful in this holistic transformation, MedTech also need to rethink their overall model, where to play, and how to win while developing new sets of capabilities for the future (such as data analytics, collective and target innovation, and digital route to market). This approach implies not just a broader scope for the transformation but also a radical change in execution: MedTech SMEs must seed and foster their capabilities and empower and mobilize the whole organization to maintain a focus on unlocking a transformation's full potential.

5.1 Limitation and Future Research

There are several limitations in this study and each of these limitations also provide opportunities for future research. Firstly, this research only focuses on MedTech SMEs in Vietnam. This is a very small industry which can lead to bias. It is recommended that future studies should cover other industries and MNCs. For example, telecommunications and agriculture which are significant sectors in Vietnam. Secondly, although the interviewees are business leaders or product development managers, eight interviews are not sufficient to cover the different perspectives on innovation. Future study should involve as many respondents as possible. And lastly, the four shifts in Vietnamese innovation have not been widely tested. Future research should empirically validate and enrich the four dimensions.

References

1. Chau, N.V.V., Dung, N.T., Geskus, R., et al.: Proactive response and innovative approaches to SARS-CoV-2 in Vietnam. BMJ Innov. **V7**, p1-6 (2021)
2. Klingler-Vidra, R., Tran, B., Uusikyla, I.: Vietnam and innovation in COVID-19 testing. BMJ Innov. **7**(Suppl 1), 19–22 (2021)
3. Thuy, N.T.T.: SME innovation management: a study of Vietnam's healthcare industry'. Thesis (Master), University of Nottingham (2018)
4. Hoang, L.: 'Asia's COVID recovery: Vietnam's breakout moment. Financial Times, January 27. https://www.ft.com/content/ffa50623-b86f-4456-a6f0-9abceeb1ef7a. Accessed on 6 June 2021
5. Galia, F., Legros, D.: Complementarities between obstacles to innovation: evidence from France. Res. Policy **33**, 1185–1199 (2004)
6. Tourigny, D., Le, C.: Impediments to innovation faced by Canadian firms. Econ. Innov. New Technol. **13**(3), 217–250 (2004)
7. Storey, J.: The management of innovation problem. Int. J. Innov. Manag. **4**(3), 347–369 (2000)
8. Freel, M.: Barriers to product innovation in small manufacturing firms. Int. Small Bus. J. **18**(2), 60–79 (2000)
9. Chung, L., Tan, K.H.: The unique Chinese innovation pathways: lessons from Chinese small and medium sized manufacturing firms. Int. J. Prod. Econ. **190**, 80–87 (2017)
10. Tan, K.H., Ji, G., Chung, L., Wang, C.H., Chiu, A., Tseng, M.L.: Riding the wave of belt and road initiative in servitization: lessons from China. Int. J. Prod. Econ. **211**, 15–21 (2019)
11. Minagawa, T., Trott, P., Hoecht, A.: Counterfeit, imitation, reverse engineering and learning: reflections from Chinese manufacturing firms. R&D Manage. **37**(5), 455–467 (2007)
12. Proctor, T., Tan, K.H., Fuse, K.: Cracking the incremental paradigm of Japanese creativity. Creativity Innov. Manage. J. **13**(4), 207–215 (2004)
13. Kim, L.: From Imitation to Innovation: Dynamics of Korea's Technological Learning. Harvard Business School Press, Cambridge, MA (1997)
14. Yin, R.K.: Case Study Research: Design and Methods, 5th edn. SAGE, London (2014)
15. Tan, K.H., Zhan, Y., Ji, G., Ye, F., Chang, C.: Harvesting big data to enhance supply chain innovation capabilities: an analytic infrastructure based on deduction graph. Int. J. Prod. Econ. **165**, 223–233 (2015)

Crowd Engineering Platform—Functions Supporting Co-creation in Product Development

Michael Hertwig, Joachim Lentes, Adrian Barwasser, and Frauke Schuseil

Abstract Digital collaboration in all sectors of industry is on the rise, also driven by the COVID-19 pandemic. People centricity, location independence and resilient delivery are important trends in this regard. Crowd engineering is a possible solution to address the challenges posed by these trends. This paper provides an insight about a technological approach for a platform for community-based product engineering. The platform consists of different cornerstones to ensure flexibility towards users, open interfaces and transparency to support re-use of existing project results. As different users have different requirements, companies have over time grown an IT infrastructure serving their needs best. Prosumers, i.e. consumers which are involved in the production of the respective products, have a limited set of digital tools allowing them to contribute to product creation. A suitable platform should be able to integrate different approaches and data from various digital tools. Rather than being monolithic, an appropriate platform needs to be a hub for core data and functionalities to support secure and reliable collaboration between all involved actors, corporate or private. These active users will be organized in a virtual community to create an attractive platform for the exchange of ideas, concepts and realizations for technical solutions by developers. The paper will introduce a possible platform structure where personal data, project related data, development data, technical product descriptions and community interactions are combined. A common access point allows an aggregation of touchpoints without centralization. Apart from the examination of technical and structural aspects of such a platform, approaches for a possible operator model are discussed. In addition to providing technically required functions, a lively community is an important basis for successful crowd engineering. Therefore, this paper will examine both community building and the implementation of co-creation projects, as they are necessary to operate the crowd engineering platform economically. Finally,

M. Hertwig (✉) · J. Lentes
Fraunhofer Institute for Industrial Engineering IAO, Nobelstrasse 12, 70569 Stuttgart, Germany
e-mail: michael.hertwig@iao.fraunhofer.de

A. Barwasser · F. Schuseil
University of Stuttgart, Institute for Human Factors and Technology Management IAT, Nobelstrasse 12, 70569 Stuttgart, Germany

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_66

different potential further advancement of the Crowd Engineering platform shall be discussed. This focuses on technological solutions that support users or serve certain tasks.

Keywords Crowd engineering · Co-creation · IT-based platform · Collaborative product development · Community-based development

1 Introduction

Product life cycles have shortened in recent years [1]. Manufacturing companies are therefore required to realize product innovations more quickly to maintain their market position. However, small and medium-sized companies in particular face special challenges here due to their limited capacities [2]. This is also accelerated by the convergence of IT-based systems [3].

At the same time, the increasing digitization of work and private life is opening up a greater need for participation. The barriers to the use of development software are falling. This is because the paradigms of Open Innovation and Web 2.0 are being applied more and more widely [4]. This makes it easier to integrate customers into the product development process. Through community-based processes, the challenges of speed are combined with the potential of integrating external stakeholders. By creating a fan community, the company gains a larger knowledge pool. This is because external participation brings fresh ideas and impulses into the processes. Crowd Engineering is a new approach to product development with enormous innovation potential.

The first step is to delineate the field of consideration by presenting the innovative approach to collaborative product development. This is followed by the presentation of a possible technical solution to support crowd engineering. The article is completed with the illumination of suitable operating concepts. Only if the platform is attractive for the community, it is available for the realization of crowd engineering projects with commercial interest.

2 Innovative Approach for Collaborative Product Engineering

2.1 Crowd Engineering

The combination of conditions of open innovation, crowd sourcing and product development lead to Crowd Engineering [5]. Crowd Engineering implies the utilization of a crowd, which represents a group of (various) participants for the planning, design and development of (technical) products and corresponding engineering activities [6]. To enable the participation of a large number of contributors all engineering data needs to be accessible by all contributors. This requires a sufficient set of IT-based tools and storage possibilities, best would be cloud-based [7].

2.2 Community

In order to flesh out a potential ecosystem, an analysis of suitable stakeholders and their interests is carried out. The application of personas allowed an extensive human-factors perspective. Based on these results, it was possible to identify different utilization scenarios.

Core results of a community development are:

- Physical events may speed up the growing of a community

  Physical event formats are particularly necessary in the start-up phase of the community. They create a sense of community and allow users to establish the necessary relationships through exchange [8]. After all, this initial community represents the nucleus for the growing community. This first group of users develops the initial identification characteristics and values of the community.
- A critical mass of participants is required to create an active community

  In order to create a longer-term active and self-perpetuating community, sufficient "noise" must be generated. This noise refers to activities on the platform that are also reflected in entries in the newsfeed and communication. In the start-up phase, this must be done by the community management. However, in order to achieve this in the long term without major efforts by community management, a critical mass of community members is required [9]. This ensures that there is sufficient activity on the platform, even if the frequency of activity varies. Based on an analysis of Balka et.al a community above 1000 members create a stable activity [10].
- Community health is dependent on regularly tasks and challenges

  Every community member has a life cycle. This life cycle varies in duration depending on the background, current life situation, interests and role of the community member. To avoid stagnation or regression after the expansion phase, the community must be continuously rejuvenated. This can be achieved by creating a regular range of opportunities for participation.
- Changes in user-interaction and tools has a negative impact on community building

  The greater the discrepancies in design, operating concept and tools used, the more difficult it is to create the impression of a consistent workspace. Different work environments can be designed differently, but it must be possible to switch between the work areas in a way that is easy to understand. This requires a single sign-on, in which all work areas are made accessible without barriers with a single log-in. The menu arrangement and design should also be similar in order to achieve a consistent operating concept.
- Strengthening the community health are essential

As in any social space, ethical hygiene plays a significant role in the community. Disruptive elements must be reduced, and positive communication strengthened. To this end, community managers must be appointed to ensure compliance with the rules in the digital space and, if necessary, to sanction any deliberate transgressions.

These core results have been qualified by a survey performed with members of open-source communities. Additional aspects have been added. Therefore, co-determination is important based on the perspective of the participants and has an impact on the path to results. For a successful community project an equal access to data and information is as important as clarified competencies and responsibilities.

3 Technological Supportive Infrastructure

As stated, a tool for interaction is required to enable the community-based processes [11]. Therefore, a digital environment was designed to support Crowd Engineering. A modular approach was chosen to support the integration of existing modules and functions [12]. This was meant to reduce doubled effort and re-development of already existing functionalities [13].

The core of the developed platform consists of three individually operatable platforms, which have been connected to fulfill all requirements of the intended core functionalities. These three core components are accessible for users by a single-sign-on solution to provide an improved user experience. The core modules are a community module, an ideation module and the Crowd Engineering project module (see Fig. 1).

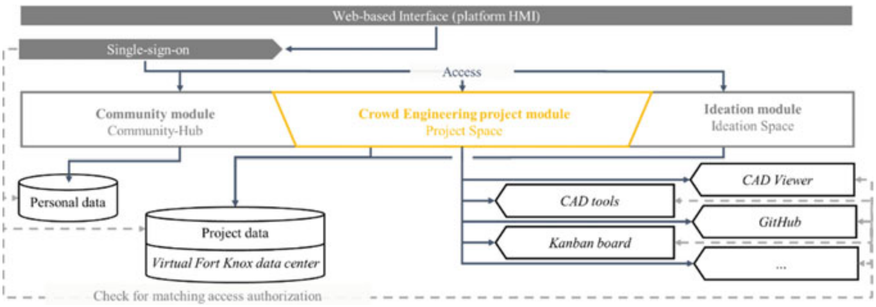


Fig. 1 Design and structuring of the crowd engineering platform [15]

3.1 Community Module

The community module is also referred to as the Community-Hub in the project context. This is where exchange, networking and social interaction take place [8]. The module is structured similarly to social networks. For this purpose, the integration of a white-label platform was carried out. It provides all the necessary functions which have been identified in different situations of a community-based project both physically operated and online supported. In this area, information about the user is accessible in the form of profile and activity information. It is also possible for users to find like-minded people, potential partners with desired skill-sets, as well as events or news in the field of interest and crowd engineering.

3.2 Ideation Module

The ideation module, which was named Ideation Space, is the functionality that represents the core function of open innovation platforms. This functionality was provided by the project partner innosabi [14]. An objective of the integrated functionality is the collection of impulses, ideas and comments from users. The aim is to open up the creativity and potential solution space. All collected impulses and contributions can be converged in later phases of the project. With the provided functionalities like collecting ideas, discussing entries, and evaluating contributions the ideation module offers all relevant functions for the digital realization of open innovation and co-creation processes. The advancement of functions were integrated in the Crowd Engineering platform, if it fitted to the defined requirements.

3.3 Crowd Engineering Project Module

Project Space became soon the name of the Crowd Engineering project module. It is the core of the platform because it enables the community-based product development. In this area the community members are to be empowered to shape the product development in crowd. It is possible to view current ongoing projects and based on the own interests and formulated needs to participate in. With participation the user get access to project data and development tools chosen by the already active community members. With these two basic aspects the users are able to contribute and add their solution approaches. All other members get information about changes performed and have the possibility to view the changes made. Pre-defined project frameworks and project process phases served as default structure for projects which could be adapted to individual preferences. The project owner is in charge to motivate the members and adapt the phases if required. Based on this, Crowd Engineering project could be performed like conventional projects but with distribute contributors with digital interactions.

4 Potential Operating Model of the Crowd Engineering Platform

The analysis of essential elements for the operation of the IT-based Crowd Engineering platform identifies three main factors—customers, contributors and facilitators (see Fig. 2).

- Customers of the platform are institutions or persons who want to develop a solution via the platform by use of Crowd Engineering. For this purpose, they provide funds or prizes for contributing users to encourage participation. This gives them the opportunity to post their tasks, challenges and needs in the community and to receive results. The exploitation of the content and results, including the scope of rights, can be linked to the amount of funds invested.
- Contributors are members of the Crowd Engineering community. The community is composed of different actors. These actors can be classified by their interests or motivation and their intensity of participation. For example, there are intrinsically motivated participants who seek to expand their own competencies, contribute ideas, or live out their desire to develop technical solutions. The focus is only secondarily on prices and remuneration. Commercial participants see the possibility of receiving financial compensation or participating in the exploitation through active collaboration in Crowd Engineering projects.
- Facilitators are required to enable the operation of the platform. There are technical tasks like maintenance, service and further development of the technical basis. This group is ensuring that the platform operates according to the requirements. At

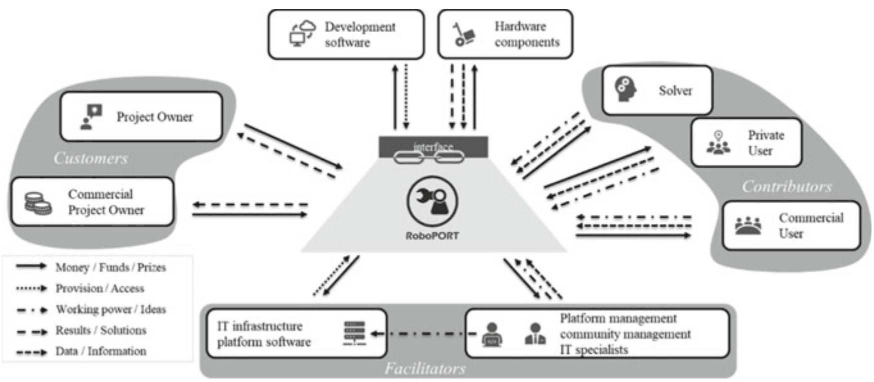


Fig. 2 Interdependencies in a potential operating model for a crowd engineering platform

the same time, there are management tasks. These tasks are subdivided into platform management, project management of commercial projects and community management.

Regarding the operating model, it becomes clear that the operator must act with two assets. The operator needs the technical platform, which provided the tools and virtual working environment. Furthermore, the community is a critical aspect for the success of community-based projects. If these are the assets, they also represent the actual cost factors. Thus, the community must be built up, matured and renewed through appropriate measures. According to the lifecycle phases of a community, there is a risk of losing members or attracting far fewer new members than are needed to have a viable and committed community in the saturation or maturation phase [16]. Accordingly, the platform operator must always strive to revitalize the existing community.

4.1 Business Model for Creating a Critical Mass of Participants on the Platform

The community-building phase is the most challenging. However, it is a crucial step to provide a benefit by Crowd Engineering. Without a reliable number of contributors, it is not possible to ensure a success for any project which shall be performed by co-creation and Crowd Engineering. Potential reasons are various, from limited number of contributors, limited experience and knowledge to loss of interest by the contributors.

In the best case a symbiotic relationship with a non-profit organization could be created. A collaboration with the University of Stuttgart was set up. By supporting teaching, the platform is used in a non-commercially oriented context. The partner of the platform operator pays a small amount to settle infrastructure costs. University professors can use the platform for their teaching and student projects. The students can use the platform to work on the projects and provide the results to the lecturer. These projects are "closed", meaning that only connected contributors have access. This ensures the privacy of the students and limits application to cheat. The results can only be viewed if both the university professor and students agree.

This is interesting for the platform operator because it creates a background noise. Thus, continuous activity exists, which contributes to a healthy community. As students become full members of the crowd engineering community, this can revitalize the community.

4.2 Business Model for Community Internal Projects (Non-commercial Projects)

In an established or mature community, community members' projects will be a core activity. The special feature here is that the community members rarely aim for a commercial benefit. The community members who are customer and developer at the same time are prosumer [17]. These persons act as project owner and are interested in the results [18]. To increase the motivation to participate in these projects, the prosumer can offer prizes, e.g. cash prizes. If this is done, the platform operator receives a small commission, which may cover infrastructure costs and support services. However, the motivation to contribute and participate can increased by other ways. Non-materialistic prices, such as community recognition are possible goals, which are aided by the establishment of a sub-community of specialized contributors, like guilds. In the case of non-commercial projects, the results are open and accessible to the community for their own use or for other projects to build on. The results are to be licensed according to open-source conditions to ensure further use [19].

For the platform, these projects are the basis for attractiveness and an active community. Since the revenues may not cover the expenses of the platform operator, this business model must be subsidized by other activities and offers.

4.3 Business Model for External Projects with Commercial Interest

To cover the expenses which are not covered by the business models discussed before, an application with the objective for profit is required. In this case, commercial users of the platform act as commercial project owners and pays for the implementation and support of the Crowd Engineering project. The money contributed is used to pay for the expenses of the platform operator as well as prize money and compensation for the contributors. To ensure the usability of the development data on the part of the commercial user, the software must comply with the appropriate licenses. Another way of remunerating the community can be through a profit-sharing scheme with correlation to the amount of contribution of each community member. The currently existing IP rights models are not yet mature, which requires individual contracts.

Another application that the platform supports is performing beta user tests to evaluate a product, including its features and functions. The commercial user triggers the provision of the product via a web store or from his own depot. The project costs serve on the one hand to cover the platform expenses (infrastructure costs and commission) and to provide the compensation of the community members. The contributors (best tester) provide their assessment, evaluation and suggestions for improvement. The commercial user receives the collected results for further use in exchange for the compensation.

5 Summary and Outlook

Crowd Engineering represent an approach to open a company's product development by integrating external participants to increase the innovativeness. It allows the competencies, opinions and experiences of relevant target groups to be integrated in the early stages of the product development process. The application of Crowd Engineering can lead to shorter development cycles and increase the customer benefit.

A core aspect is the creation of a digital working environment which supports the collaboration beyond any single company. A potential structure of an IT-based platform was presented and discussed. Furthermore, required operating models have been presented. These were discussed regarding their intension and influence on the performance of Crowd Engineering. This provides a base for introducing community-based product development industrial processes.

The subject of ongoing research is the evaluation of the required business models to ensure a reliable operating concept. Aspects of need for further research are consideration of intellectual property of each contributor. This includes different aspects like evaluating the quality and amount of contribution and corresponding interdependencies for developments build on that. On the other hand, the participation based on commercial benefit based on cross-organizational developments of many is still not easily possible which leads either licensing or publishing the results as open-source.

Acknowledgements The authors would like to thank the sponsors and partners of the project "RoboPORT" without whom this publication could not be accomplished. First, the PAiCE sponsorship from the German Federal Ministry for Economic Affairs and Energy (BMWi) for its financial support (Grant no. MA16003). Furthermore, the whole project team; especially all project members from innosabi GmbH, General Interfaces GmbH, BSH Hausgeräte GmbH and UnternehmerTUM GmbH.

References

1. Deutscher Bundestag: Zur Diskussion um die Verkürzung von Produktlebenszyklen. Dokumentation (WD 5-3000-053/16) (2016)
2. Yadollahi Farsi, J., Toghraee, M.T.: Identification the main challenges of small and medium sized enterprises in exploiting of innovative opportunities (case study: Iran SMEs). J. Glob. Entrepreneurship Res. **4** (2014)
3. Andersen, N., Börsch, A., Blohmke, J.: Innovation in Small and Medium Enterprises. How Startup Management Practices Can Help in Solving Digitalization Challenges. Munich (2017)
4. Aquilani, B., Piccarozzi, M., Abbate, T., Codini, A.: The role of open innovation and value co-creation in the challenging transition from industry 4.0 to society 5.0: toward a theoretical framework. Sustainability (2020). https://doi.org/10.3390/su12218943
5. Panchal, J.H.: Using crowds in engineering design—towards a holistic framework. In: International Conference on Engineering Design, ICED15. Politecnico Di Milano, Italy (2015)
6. Adam, F., Hertwig, M., Barwasser, A., Lentes, J., Zimmermann, N., Siee, M.: Crowd engineering—approach for smart and agile product development in networks. Proc. Manuf. **39** (2019)

7. Eigner, M., Eiden, A. Apostolov, H.: Crowd-engineering—bringing full cloud cad into the lab. In: International Conference on Engineering and Product Design Education (2017)
8. Chroneer, D., Johansson, J., Nilsson, M., Runardotter, M.: collaborative network for industry, manufacturing, business and logistics in Europe. In: Requirements for Business Models and Collaboration Patterns in Supply Chains, D1.2 (2017)
9. Seran, S., Izvercian, M.: Prosumer engagement in innovation strategies. Manag. Decis. (2014). https://doi.org/10.1108/MD-06-2013-0347
10. Balka, K., Raasch, C., Herstatt, C.: The effect of selective openness on value creation in user innovation communities. J. Prod. Innov. Manag. (2014). https://doi.org/10.1111/jpim.12102
11. Chai, Y., Miao, C., Sun, B., Zheng, Y., Li, Q.: Crowd science and engineering: concept and research framework. Int. J. Crowd Sci. (2017). https://doi.org/10.1108/IJCS-01-2017-0004
12. Stol, K.-J., Fitzgerald, B.: Two's company, three's a crowd: a case study of crowdsourcing software development. In: Jalote, P., Briand, L., van der Hoek, A. (eds.) Proceedings of the 36th International Conference on Software Engineering. ICSE '14: 36th International Conference on Software Engineering, Hyderabad India, pp. 187–198. ACM, New York, NY, USA (2014). https://doi.org/10.1145/2568225.2568249
13. Calcott, G.: Microservices versus Monolithische Architekturen: ein Leitfaden | silicon.de. https://www.silicon.de/41666855/microservices-vs-monolithische-architekturen-ein-leitfaden (2018). Accessed 21 Jan 2021
14. van Delden, C., Chia, D.: Connect the Dots. Agile Innovation and Collaborative Ecosystems, 1st edn. München (2020)
15. RoboPORT-Consortium: Crowd Engineering. Ein neuer partizipativer Ansatz, (Robotik-) Produkte zu entwickeln und zu realisieren. Vorstellung von Ergebnissen und Erkenntnissen im Projekt RoboPORT—Robotic Platform and Open Repository for Toolkits, Stuttgart (2020)
16. Bullinger, H., Baumann, T., Fröschle, N., Mack, O., Trunzer, T., Waltert, J.: Business communities. In: Professionelles Beziehungsmanagement von Kunden, Mitarbeitern und B2B-Partnern im Internet, 1st edn. Galileo Press, Bonn (2002)
17. Lan, J., Ma, Y., Zhu, D., Mangalagiu, D., Thornton, T.: Enabling value co-creation in the sharing economy: the case of Mobike. Sustainability (2017). https://doi.org/10.3390/su9091504
18. Humphreys, A., Grayson, K.: The intersecting roles of consumer and producer: a critical perspective on co-production, co-creation and prosumption. Sociol. Compass (2008). https://doi.org/10.1111/j.1751-9020.2008.00112.x
19. Dopfner, M., Weitzl, V.: Die Entwicklung von Open Source und die Disruption der Hardware-Industrie | Digital Society Blog. https://www.hiig.de/die-entwicklung-von-open-source-und-die-disruption-der-hardware-industrie/ (2017). Accessed 21 Jan 2021

Potential Contributions of Artificial Intelligence in Crowd Engineering

Frauke Schuseil, Joachim Lentes, Michael Hertwig, and Adrian Barwasser

Abstract Opening up product development in companies to external interdisciplinary participants offers the opportunity to develop products faster and more cost-effectively. An early integration of potential customers in the product development process enables a stronger focus on their needs and thus facilitates customised product development. A lack of suitable tools and rigid processes within companies, however, make this integration unpractical in many cases. In this context, Crowd Engineering allows the involvement of an engineering community, with community members being integrated in such a way that they can participate actively and in a distributed manner in product development using a central internet platform. Meanwhile, artificial intelligence methods and approaches are becoming more sophisticated and acquire more importance in various aspects of private life and business context by providing new solutions to topics which formerly were not solvable or only with high efforts. This paper analyses how artificial intelligence can support product development, especially in the context of Crowd Engineering. It thereby provides an overview of how artificial intelligence already supports development processes today and which potentials arise from this. Furthermore, an analysis of the collaboration between humans and artificial intelligence is carried out in the context of Crowd Engineering. The main research questions are how artificial intelligence can support in collaboration and how the role of humans is being shaped and transformed. Another key question is how collaboration can be achieved in the context of Crowd Engineering—hence this paper proposes collaboration modes that vary according to three dimensions of collaboration: degree of AI-assistance, degree of task complexity and the degree of AI-autonomy.

F. Schuseil (✉) · A. Barwasser
University of Stuttgart IAT, Institute of Human Factors and Technology Management, Nobelstraße 12, 70569 Stuttgart, Germany
e-mail: frauke.schuseil@iat.uni-stuttgart.de

J. Lentes · M. Hertwig
Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering IAO, Nobelstraße 12, 70569 Stuttgart, Germany

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_67

Keywords Crowd engineering · Artificial intelligence application · Collaboration · Product development · Open Innovation

1 Introduction

Crowd Engineering is a product development approach, which harnesses an interdisciplinary group of individuals to participate in engineering activities like e.g., planning, design, and development of products [1]. When external participants are involved in the product development process, this can, among others, result in the following benefits:

Reduced development time, increased innovation capability, and higher customer acceptance [1]. Since this approach is based on the concept of involving distributed participants from the community, an IT-supported platform is required as a basis (Fig. 1). Artificial Intelligence (AI) technologies and methods have improved rapidly in the past few years, which enables high-tech applications with ever improving results [2]. Therefore, AI can be applied to facilitate collaboration in Crowd Engineering, so that the crowd participants themselves have more time to focus on purely human work, can be relieved of highly repetitive work, or to expand their skills and abilities [3]. Such collaboration offers many potentials but also brings challenges, as machines do not think like humans and require to be controlled appropriately. This article is structured as follows: In the next chapter, a brief overview of product development already supported by AI is given and potentials as well as challenges for its use in Crowd Engineering are illustrated. In Sect. 3, the collaboration between humans and AI is examined and a collaboration model is proposed. Furthermore, the role of humans in this context is analysed. Finally, a conclusion and an outlook are provided.

2 Using AI for Crowd Engineering

In the context of this paper, Buchkremer's [4] definition of Artificial Intelligence is used, according to in which AI includes the following:

- use and analysis of the application of human senses (reading and writing, seeing, feeling, hearing, speaking)
- as well as human understanding, learning, adaptation and reasoning,
- with non-human means, such as a computer.

Pan [5] presented an outlook on AI 2.0 in 2016, noting that large communities show extraordinary intelligence through participation and interactions of individuals over the internet, forming a new type of intelligent system. This chapter reviews how

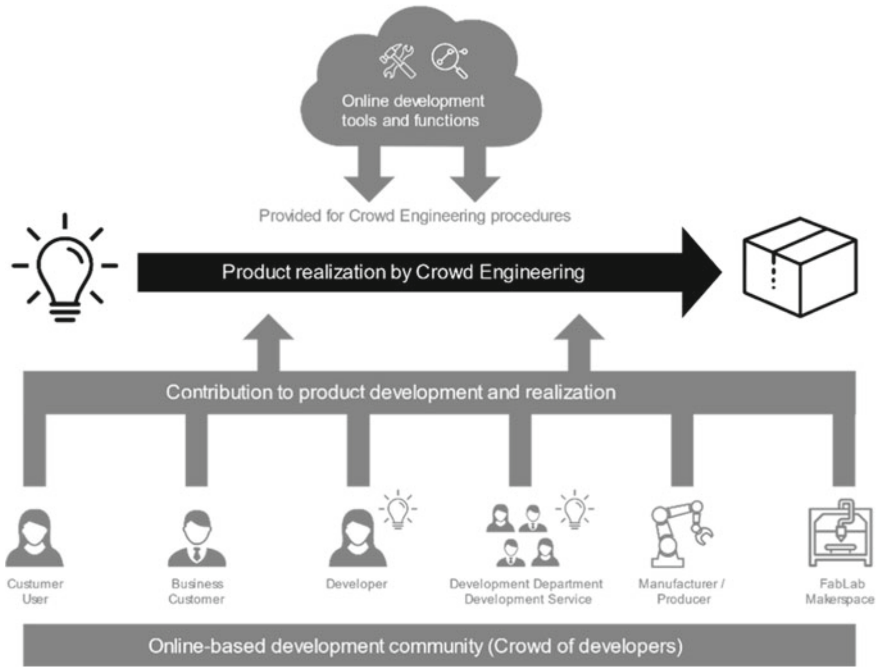


Fig. 1 Overview over crowd engineering with respective contributors

artificial intelligence is already supporting product development today and identifies potentials and challenges for the application of AI in the context of Crowd Engineering.

2.1 Product Development with Artificial Intelligence

Nowadays, product development is already permeated by diverse AI approaches. One example is the use of AI for static mechanic finite-element simulations. There are approaches to support the pre-processing procedure with neural networks, classification and regression methods or plausibility checks of static mechanic FE simulations with convolutional neural networks [6]. In general, ML (Machine Learning), NLP (Natural Language Processing) or ontological approaches are nowadays used frequently to support engineering purposes [7–10].

The use of ontologies is also suitable for the application in the field of product development. For instance, one approach is the utilisation of ontologies for knowledge representation, for example an ontology platform for distributing information along the product life cycle [11] or the development of an ontology-based product design framework for manufacturability verification and knowledge reuse [12].

Besides, Machine Learning is a good enhancement for applications with digital twins in engineering, e.g., in the context of simulation [13].

2.2 Potentials and Challenges of Using AI for Crowd Engineering

Managing Crowd Engineering with real humans as participants is one thing—also managing collaboration with machines on the other hand is a demanding challenge. There are many differences between machines and humans to be considered, an important one is the fact that humans have a consciousness—and computers don't, which means that they act upon defined objectives only [3]. However, if an AI is to be useful, it must be able to do more than just support with tasks; for example, it needs to be suitable for problem solving and critical thinking, to propose solutions or make decisions on its own [14]. Collaboration between humans and AI can add value provided that the different partners support each other and do not compete against each other (Fig. 2). It is important that each participant can build on its strengths: While a human can think abstractly and see the larger context, as well as have intuition, creativity, empathy and social skills, an AI can handle large amounts of data and repetitive tasks and scores high on speed and scalability [3, 15]. The result of ideal collaboration between humans and machines then promises better results than the best performances of the partners alone [3, 16]. But how can we ensure optimal collaboration and cooperation between these very different partners? Many challenges are to be considered—how can the interaction between humans in the crowd and AI be organised, how are ethical and moral aspects considered and which tasks are left for the human in the product development process?

3 Collaboration

A machine that can act as a team member is an autonomous, proactive, and sophisticated technology that can make inferences and gain new insights from information, learn from past experiences, contrast consequences of potential choices, propose solutions, and participate in decision-making with humans [14]. To enable collaboration between humans and such machines raises issues that are to be considered.

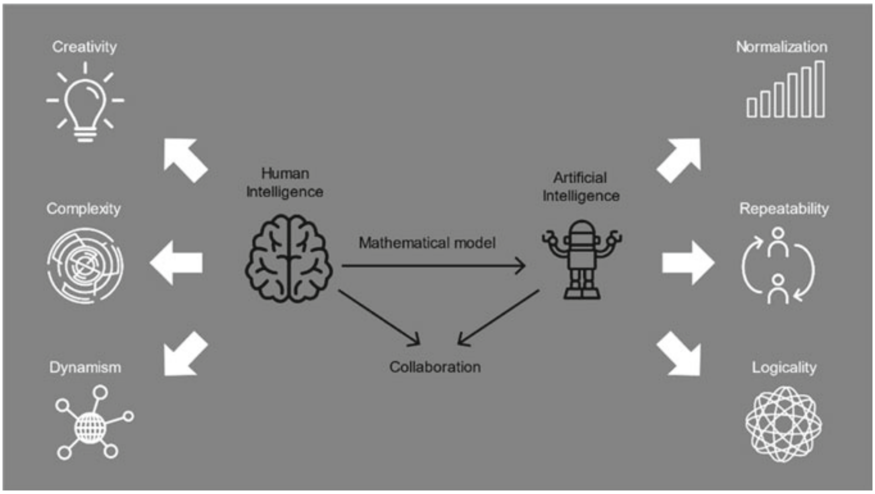


Fig. 2 Difference between human and artificial intelligence and respective selected strengths according to [17]

3.1 Human-AI Collaboration

Three dimensions of collaboration between humans and AI are presented here: degree of AI-assistance, degree of task complexity and the degree of AI-autonomy. They are summarised in a matrix (Fig. 3), from which four collaboration modes can be derived.

1. *Degree of AI-assistance:*
   The collaboration between humans and AI, or the degree of AI-assistance, can range between two endpoints: on the one hand, machine support is limited, and humans are left to do most of the work themselves, however, machines support human decision-making. They can amplify cognitive abilities, embody human capabilities to augment our physical skills and interact with persons so that one can concentrate on higher-value tasks [3, 16]. At the other end, the AI takes over many tasks itself and is thus very independent, the level of support is therefore high, and the human supports the AI. Sensitive tasks depend on human involvement. They train machines to execute special tasks, explain the corresponding results and sustain responsible handling of machines [3, 16]. To enable equal collaboration in Crowd Engineering without one of the partners taking the lead, a good level can be found lying around midpoint, so that the AI can take on many tasks without the human always having to assign them and in such a way that the human is not "commanded" by the AI. The optimal situation is when both partners take on tasks independently and work together on an equal footing.

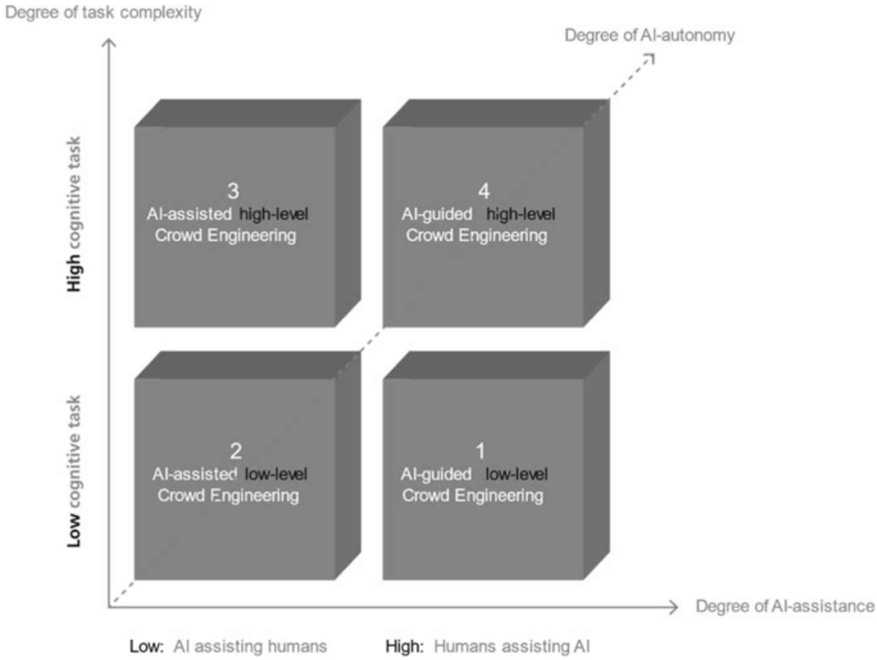


Fig. 3 Human-AI collaboration matrix with three features: (1) degree of AI-assistance, (2) degree of task complexity, and (3) degree of AI-autonomy

2. *Degree of task complexity:*
   The dimension task complexity varies on the one hand between tasks that are more suitable for machines, e.g., logical tasks or repetitive tasks, i.e., those tasks that can be performed more efficiently by machines. Thus, they are of a low cognitive nature, which does not mean that they are not cognitively demanding, but that they are rather machine-processable. On the other hand, there are tasks of a human character, such as creativity, social tasks, etc. (see Sect. 2.2), i.e., jobs where human cognitive abilities are an advantage. Thus, these tasks are of a high cognitive nature. Consequently, collaboration modes in Crowd Engineering are favourable, where the human partner recognises and selects (or is assigned) such tasks that are in the upper half of the matrix and the AI selects the lower tasks in contrast.
3. *Degree of AI-autonomy:*
   In this dimension, the degree of collaboration varies from AI assisting humans with guidance to AI making own decisions. Collaboration between humans and machines can be realised between two extremes: decisions made solely by humans and autonomous decision-making by the AI with the relevant collaboration forms ranging between these two, such as assistance in decision-making as well as verification [15]. The degree to be chosen depends on the task, individual preference, and context (e.g., private, business, etc.) [15].

Realising Crowd Engineering in the context of human-AI collaboration:

The characteristics of the human-AI collaboration matrix (above dimensions: degree of AI-assistance, degree of task complexity and degree of AI-autonomy) lead to four different potential applications of AI in Crowd Engineering, depending on the respective degrees of the three levels:

1. **AI-guided low-level Crowd Engineering**: The first case of human-AI collaborative engineering is focused on machine-suitable (engineering) tasks controlled by AI, e.g., project planning or task distribution.
2. **AI-assisted low-level Crowd Engineering**: The second case involves (engineering) tasks that are most likely to be suitable for machines, but where the machine is only a supporting agent and the human both guides and decides. An example of such collaboration in the context of Crowd Engineering is that the AI calculates and suggests optimisations during construction tasks.
3. **AI-assisted high-level Crowd Engineering**: The third case refers to demanding cognitive tasks in which AI provides support. One example is the suggestion of ideas or design proposals.
4. **AI-guided high-level Crowd Engineering**: The last case of human-AI collaborative engineering relates to cognitive demanding tasks that are guided and autonomously implemented by the AI. The AI makes the decisions (e.g., decision for a design proposal or similar). This level thus describes the most autonomous manifestation of the matrix.

3.2 Role of Humans

Crowd Engineering thrives on participants within the community interacting and collaborating with each other. When machines become a part of it, the role of humans must adapt in the development process so that both sides can benefit. Humans should therefore concentrate on *human* tasks requiring creativity, critical thinking, social and emotional skills [3]. For successful teamwork, a person must be willing and able to cooperate with a machine. By accepting AI as a useful and valuable addition to the team, many tasks can be done faster and more efficiently. Most administrative tasks, for example, as well as product optimisation in the case of engineering can be taken over by machines. In the case of AI-guided high-level Crowd Engineering, it must be determined whether the human participants in Crowd Engineering are informed that they are working together with machines. One risk in this type of interaction is that of misunderstandings. As of today, most people have little experience working with AI in this context. It would be negative if the AI cannot determine that it knows too little about a particular area and would therefore need human expertise, but this demand is not discovered.

In a study on interactive collaboration between humans and AI, researchers at Seoul National University [18] came to the following conclusions, including:

- People prefer to receive detailed information/instructions when communicating with an AI (as opposed to simple instructions). The experience was best when just the right amount of instructions were provided (not too little, but not too many instructions either).
- It is preferred to be able to take the initiative and have the AI provide detailed explanations (but only if requested). It is important to note that users always wanted to make the decision (as a differentiation from the AI). The AI, on the other hand, should take over tedious and tiresome tasks.
- AI can reduce the user-perceived predictability, understandability, and controllability of tasks, while detailed instructions can offset these negative effects. Moreover, low predictability can even increase user enjoyment. It became apparent that AI may well be seen as a human partner, but it was always considered subordinate to humans.

Although the study addressed co-creation with AI in the context of drawing, various learnings and best practices can be derived from this for AI-supported Crowd Engineering, because both show a collaboration mode that includes creativity and cooperation:

1. Human act as decision-maker, AI as facilitator. AI can make suggestions, but the final decision should be left to the human, if only to prevent misunderstandings or misjudgements by the AI.
2. AI must work transparently—it must be obvious whether it is working “human-like” or “machine-like”, as well as the work process and how it comes to conclusions or decisions.
3. When tasks are defined by the AI, the task description must be to the right extent (not too little, but also not too detailed and only if required).

Additionally, people will prefer to cooperate with an AI if this interaction is fun and “works” [18], which leads to the realisation that an AI is to be designed in such a way that it eases the work and makes cooperation enjoyable.

For an AI to be a useful team member in comparison and to support humans, its role must also adapt to the requirements of a complex working environment in distributed Crowd Engineering: it must perform some of the steps of the problem-solving process, e.g., define problems, identify causes, propose and evaluate solutions, choose between options, create plans, take actions, learn from past interactions and participate in debriefings [14]. When AI joins the team, it can become an equal partner in a hybrid human-AI environment [15]. For this to work, the AI must especially learn and maintain moral principles [15]. An important aspect in the context of AI-supported Crowd Engineering is the understandability and comprehensibility of AI algorithms. Particularly in cases where the AI controls activities in the process, human team members expect the AI to act transparently. This is especially important because Crowd Engineering is based on partnership-based collaboration and this can only work if the partners trust each other. This trust can only be built if the person giving it is given a comfortable level of model understanding and transparency but should always take into account both algorithmic properties and human needs [19].

4 Conclusion and Outlook

In this paper, potential contributions of Artificial Intelligence and its collaboration modes with humans in the context of Crowd Engineering were studied. Since Crowd Engineering enables distributed and interdisciplinary collaboration, it seems reasonable to consider AI-support as another crowd participant. Here, a collaboration model was proposed, divided into three dimensions, degree of AI-assistance, degree of task complexity, and degree of AI-autonomy, which allows four corresponding Crowd Engineering models. This research is not exhaustive, as further studies are needed to address certain remaining questions. Above all, it would be important to conduct a trial of how Crowd Engineering works together with AI-participants and to draw appropriate conclusions. The theoretically elaborated aspects of the collaboration of humans and AI in Crowd Engineering will have to be validated in concrete use-cases. Further topics to be addressed include confidence in the AI, what decision-making should look like in such a constellation, and the extent to which an AI must work transparently and comprehensibly.

References

1. Adam, F., Hertwig, M., Barwasser, A., Lentes, J., Zimmermann, N., Siee, M.: Crowd engineering—approach for smart and agile product development in networks. Proc. Manuf. **39**, 1317–1326 (2019)
2. Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., Malhotra, S.: Notes from the AI frontier: insights from hundreds of use cases. McKinsey Global Institute (2018)
3. Woo, W.: Future trends in I&M: human-machine co-creation in the rise of AI. IEEE Instrum. Meas. Mag. **23**(2), 71–73 (2020)
4. Buchkremer, R.: Natural language processing in der KI. In: Künstliche Intelligenz in Wirtschaft & Gesellschaft, pp. 29–45. Springer Gabler, Wiesbaden (2020)
5. Pan, Y.: Heading toward artificial intelligence 2.0. Engineering **2**(4), 409–413 (2016)
6. Bickel, S., Spruegel, T.C., Schleich, B., Wartzack, S.: How do digital engineering and included AI based assistance tools change the product development process and the involved engineers. Proc. Des. Soc. Int. Conf. Eng. Des. **1**(1), 2567–2576 (2019)
7. Fougères, A., Ostrosi, E.: Intelligent requirements engineering from natural language and their chaining toward CAD models. TMCE 2020 (2020)
8. Colombo, G., Mosca, A., Sartori, F.: Towards the design of intelligent CAD systems: an ontological approach. Adv. Eng. Inform. **21**(2), 153–168 (2007)
9. Han, J., Sarica, S., Shi, F., Luo, J.: Semantic networks for engineering design: a survey (2020)
10. Krahe, C., Iberl, M., Jacob, A., Lanza, G.: AI-based computer aided engineering for automated product design—a first approach with a multi-view based classification. Proc. CIRP **86**, 104–109 (2019)
11. Lentes, J., Zimmermann, N.: amePLM: a platform providing information provision in engineering. Int. J. Prod. Res. **55**(13), 3832–3841 (2017)
12. Li, Z., Zhou, X., Wang, W.M., Huang, G., Tian, Z., Huang, S.: An ontology-based product design framework for manufacturability verification and knowledge reuse. Int. J. Adv. Manuf. Technol. **99**(9), 2121–2135 (2018)
13. Madni, A., Madni, C., Lucero, S.: Leveraging digital twin technology in model-based systems engineering. Systems **7**(1), 7 (2019)

14. Seeber, I., Bittner, E., Briggs, R.O., De Vreede, T., De Vreede, G.J., Elkins, A., Maier, R., Merz, A., Oeste-Reiß, S., Randrup, N., Schwabe, G., Söllner, M.: Machines as teammates: a research agenda on AI in team collaboration. Inf. Manage. **57**(2), 103174 (2020)
15. Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., Söllner, M.: AI-based digital assistants. Bus. Inf. Syst. Eng. **61**(4), 535–544 (2019)
16. Wilson, H., Daugherty, P.: Collaborative intelligence: humans and AI are joining forces. Harv. Bus. Rev. **96**(4), 114–123 (2018)
17. Zheng, N.N., Liu, Z.Y., Ren, P.J., Ma, Y.Q., Chen, S.T., Yu, S.Y., Xue, J., Chen, B., Wang, F.Y.: Hybrid-augmented intelligence: collaboration and cognition. Front. Inf. Technol. Electron. Eng. **18**(2), 153–179 (2017)
18. Oh, C., Song, J., Choi, J., Kim, S., Lee, S., Suh, B.: I lead, you help but only with enough details: understanding user experience of co-creation with artificial intelligence. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2018)
19. Zhu, J., Liapis, A., Risi, S., Bidarra, R., Youngblood, G.M.: Explainable AI for designers: a human-centered perspective on mixed-initiative co-creation. In: 2018 IEEE Conference on Computational Intelligence and Games (CIG), pp. 1–8 (2018)

Framework for the Identification of Fields of Innovation in the Product Environment Via Text Mining and Semantic Networks

Michael Riesener, Maximilian Kuhn, Hendrik Lauf, and Günther Schuh

Abstract The innovation process of manufacturing companies is characterized by uncertainty caused by various external influences. Today, companies are not able to handle the amount of data available for the identification of changes within their products' environment although, this data can give valuable indications for the urgency or possibility to innovate their products. Approaches from environmental scanning, text mining and semantic networks have the potential to address individual aspects of that problem separately but there is no method for monitoring the environment holistically. The presented framework closes this gap by combining methods from the three areas that complement each other for the identification of potential fields of innovation for existing products. The aim of this paper is to enable companies to process text data from external data sources automatically and therefore generate insights for the detection of potential fields of innovation within the environment of their products. For this purpose, the product is described as a semantic network. Further, relevant external influences, i.e. customers and competing companies or respective external data sources for their description are identified. A text mining approach extracts the topics covered by these data sources and expands the product specific semantic network by linking them based on their co-occurrence with each other and the product description. Finally, the potential fields of innovation are identified and evaluated in terms of external relevance and the company's competence.

Keywords Innovation management · Innovation mining · Text mining · Topic modeling · Semantic network

1 Motivation

The development of innovative products is the core business activity of manufacturing companies. However, the transformation from a traditional seller's market to

M. Riesener · M. Kuhn · H. Lauf (✉) · G. Schuh
Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany
e-mail: H.Lauf@wzl.rwth-aachen.de

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_68

a buyer's market requires to take various external influences such as society, environment, market or customers into account [1]. Product development results can be improved, for example, by a high customer involvement [2]. The consideration of these influences are crucial as an innovation is not limited to the degree of novelty; instead, an innovation is an invention that becomes commercially successful. With regard to the degree of innovation, a distinction can be made between incremental innovations and radical or disruptive innovations [3].

In terms of product innovation, companies rate "sensing new technologies and trends" as a key success factor. At the same time, this point is also seen as a key obstacle [4]. Therefore, the detection and interpretation of indicators in their environment for the urgency or possibility to innovate their products causes uncertainty among companies [5]. Information acquisition and analysis is a possible option to reduce this uncertainty as 80% of all innovations are a recombination of existing knowledge [6].

Furthermore, 80% of all data available in business context is text data [7]. However, the analysis, evaluation and interpretation of this text data are associated with a high personnel effort [8]. Especially small and medium-sized enterprises cannot make use of analyses and reports due to a lack of personnel capacity [9]. Neglecting information risks missing trends or technical advancements [8]. A solution for the reduction of required effort is the application of text mining, but only 10% of companies use data mining and big data to systematically monitor their own business and the competition. One reason for this is that sufficient tools for the environment analysis are currently not available [10], although some use cases in that field demonstrate the potential of data and text mining techniques for product development [11].

In addition, successful innovations require more than information gathering and analysis. They are often derived from analogies, where patterns from known fields are transferred to new fields [12]. Therefore, the identification of innovation opportunities requires creativity and the ability to think in networked systems [1]. Hence, the product as well as product environment must be modeled accordingly in order to be able to depict interrelationships. Semantic networks map information in a way that is easily accessible to both humans and machines [13].

The following framework addresses this topic by presenting a procedure model that defines the consecutive steps, which are required for the implementation of the identification of potential fields of innovation by the aid of text mining and semantic networks in the product environment.

2 Relevant Terminology, Techniques and State of the Art

Based on the described motivation, there are three main challenges that have to be addressed in this context. External influences on product development have to be identified and allocated with corresponding data sources for the gathering of information about the product environment. The text data within these data sources has to be processed automatically to make the method applicable for a company independent from its size and capacity. For the identification of potential innovations within the environment of the respective product, the information contained in the data sources have to be linked to that product. Therefore, three main research areas are relevant for the solution of these challenges: environmental scanning, text mining and semantic networks, whose terminology and techniques are presented and critically reflected below.

2.1 Environmental Scanning

Environmental scanning, according to Albright, is a method of including information about external influences, such as emerging issues or trends that may affect the company's future, into internal decision-making processes and corporate direction [14]. Horizon Scanning is a subset of environmental scanning and is particularly focused on the future [15] as it enables the "systematic search for incipient trends, opportunities and constraints that might affect the probability of achieving management goals and objectives" [16]. In particular, attention is paid to the identification of so-called weak signals in order to reduce uncertainty [5] and to identify emerging issues at an early stage [17]. The term weak signal can be understood as "premature and imperfect information […] which indicate the incoming of discrete shocks or new developments in powerful trends" [18]. However, due to the large amount of data, the environmental scanning and search for weak signals fail mainly because of the high manual effort as well as the subjectivity of the procedure [8, 14, 19]. Ernstsen et al. transferred the approach of horizon scanning onto the innovation process and defined how the technique may be used for the identification and monitoring of technologies for the disruptive innovation [20]. The environmental scanning demonstrates the potential of a holistic differentiated observation of external influences, especially for product development, although the presented approaches are not automated yet.

2.2 Text Mining

In general, data mining is a method of identifying generally valid, non-trivial, new, useful, understandable patterns as knowledge from data [21]. Text mining is a special form of data mining and differs in the underlying database, which has to be text data [22]. Text data is unstructured data consisting of strings or words [23]. For the effective application of a text mining method, the application of a suitable data preprocessing is necessary, which derives structured data for the application of classical data mining applications from the naturally occurring unstructured raw data by means of a series of selective methods [24]. One use case for the application of text mining is the extraction of topics from the available documents (topic modeling) by a generative probability model that interprets topics as probability distributions over the occurring words [25]. One method that can be used for probability-based topic modeling is Latent Dirichlet Allocation (LDA) developed by Blei et al. [26]. As an exemplary application of text mining in this field, Yoon and Griol-Barres et al. identify relevant trends by analyzing the occurrence of keywords in documents and categorize them as the previously mentioned weak signals [19, 27]. Kölbl et al. identify new trends for product development by analyzing large amounts of text data by the aid of LDA [8]. These approaches show that text mining has the potential to enhance the databased product development, but up to this point, their use is limited to specific data sources or text types. They also do not interconnect trends with each other and show their relationships nor do they show the trends' impact on an existing product.

2.3 Semantic Networks

The term "semantic network" originates from the linguist Quillian [28]. Semantic networks play a central role in knowledge representation [29], because the stored information is not only readable for humans but also machines [13]. A semantic network is a "graph structure for representing knowledge in patterns of interconnected nodes and arcs" [30]. According to Brachman as well as Giarratano and Riley, knowledge can only be stored and retrieved in the semantic network by connecting the nodes and thus the contained terms [31, 32]. There is no uniform structure of semantic networks [33]; instead, they are designed according to the application. Ahmed et al. present an approach for the development of a semantic network for the engineering design discipline for either knowledge management or the representation of a common design language [34]. Shi et al. introduce a semantic network that serves as a common information source for various design and engineering purposes across different companies by analyzing data via text mining [35]. Although, semantic networks have the potential of enhancing the information retrieval, their use in product development is limited. Especially, the derivation of product-specific semantic networks for modeling the external influences on the product is not addressed yet.

3 Framework

As shown in the previous review of the existing approaches, environmental scanning, text mining or semantic network methods have the potential to solve parts of the depicted issue individually. However, there is no integrated method, which addresses this issue holistically. Therefore, the key research question for the this paper can be derived as follows:

> *How can relevant innovation potentials in the product environment be identified with the help of a product-specific semantic network and the application of text mining?*

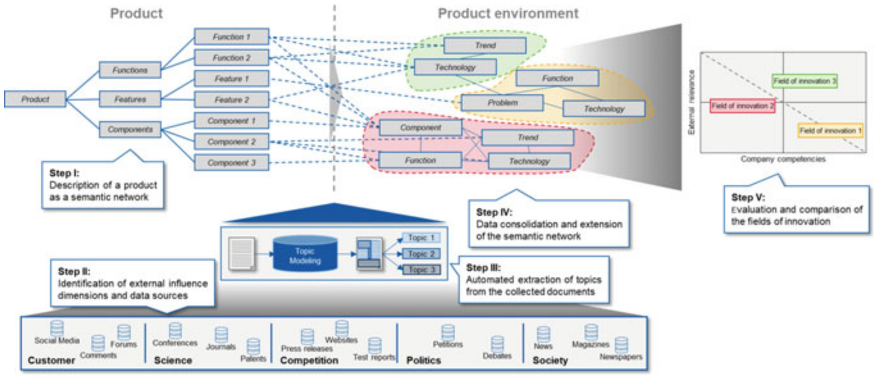


Fig. 1 Overview of the framework for the contextual identification of fields of innovation via text mining and semantic networks

Based on these core challenges, a five-step framework was developed to enable the product-specific identification of potential fields of innovation. This framework is shown in Fig. 1.

The first step is the description of an analyzed product in form of a semantic network that models all relevant product aspects and their interrelationships. For the data retrieval, the external influences on the product development that should be considered during the analysis with corresponding data sources have to be identified. Afterwards, the text data contained in the data sources must be analyzed automatically using a topic modeling approach. The identified topics are transferred into a semantic network for describing the product environment, which complements the initial semantic network for the product description. Finally, potential fields of innovation are identified within the network.

These five steps of the presented framework are described in detail below.

3.1 Semantic Description of a Product

The first step for the identification of product-specific fields of innovation is the qualitative description of product aspects that are relevant for the incremental innovation. Therefore, product aspects that a company actively designs during product development to address external influences or to advance the product are considered in this context. Such aspects may be the product's functions, features or physical components.

After the relevant product aspects are identified, they must be described semantically by means of thematically oriented nodes and the modeling of the dependence of these nodes by appropriate edges. Each node represents for example a specific function, feature or component of the product.

For modeling the product according to the above specifications, a company has three types of internal and external information sources at its disposal, which are addressed using different approaches. The implicit knowledge about the product that employees of the company possess must be transferred into a semantic structure. Other information sources are unstructured and semi-structured internal company data. In particular, semi-structured data (such as parts lists, folder directories or configurators) offer the advantage that the relations necessary for the modeling of the semantic network are already implicitly contained in the data. Additional information for the construction of the semantic network are stored in company-external structured data sources such as WikiData or WordNet, from which the nodes and edges relevant for the description of a specific product can be extracted.

3.2 Identification of External Influences and Need for Information

In the product's environment, continuous changes or signals occur from which the necessity or possibility for innovation of the product arises. Companies have to detect and interpret these signals in order to derive the right conclusions for the design of the product's aspects. For this purpose a company has to consider various external influences (e.g. society, environment, culture, customer, market, resources) [1] to innovate its product accordingly. Hence, the second step of this framework has to be the identification of the various external influence dimensions within the product environment that should be taken into account.

For deriving fields of innovation from the analysis of these external influence dimensions, external data sources are assigned to each influence dimension. For example, social networks can be used for analyzing the customer or scientific publications for science. With regard to the subsequent procedure, only data sources for text data should be considered because the following steps will apply an automated text mining approach for the data analysis.

3.3 Topic Modeling

For the analysis of the external dimensions of influence, large amounts of data have to be processed and interpreted. A manual analysis of the data is not practicable due the high manual effort. For this reason, the framework applies a text mining procedure for the automated analysis of the data.

For this purpose, a corresponding preprocessing of the data has to be conceptualized, which is mainly responsible for the transformation of text data from human-readable to machine-readable data. The targeted data format in this context is a bag-of-words. Typical data preprocessing steps are normalization, tokenization, n-gram identification, stop word elimination and noise removal [36].

After the data has been prepared accordingly, it can be evaluated with the text mining algorithm. For this application, a topic modeling method is implemented for the extraction of those topics that are addressed by the respective documents in particular. A suitable method for this use case is Latent Dirichlet Allocation (LDA), which was developed by Blei et al. [26], a generative probability model that allocates each document with the topics that it covers most probably.

3.4 Semantic Modeling

Subsequently, the recorded information must be condensed and linked to the product to model the product environment on the one hand and the influence of the environment on the product on the other. For this purpose, the topics of the analyzed documents are linked with the semantic network of the product.

According to Reichenberger, the principle of object identity applies to semantic networks. According to this principle objects should only occur once within the semantic network and all information about this object is bundled at the respective position [13]. However, the objects of the network presented here are determined empirically and independently. For this reason, the objects of the semantic network have to be checked with respect to their similarity in content. Thus, all objects can be checked in pairs regarding their semantic similarity by using a neutral database (e.g. WordNet) [37].

Subsequently, the semantic network links the topics to each other as well as to the initially defined network for the product description on the base of their co-occurrence, i.e. their common occurrence in all documents, to form a semantic network. In addition, the respective topics can be evaluated according to their occurrence in the respective dimensions of influence that were identified previously in the product environment. In this way, the resulting semantic network allows a differentiated view on topics that have an impact on product development through different dimensions of influence.

3.5 Derivation of Potential Fields of Innovation

Finally, the methodology has to enable the use of the semantic network for the derivation of suitable fields of innovation of the product. For this purpose, the topics contained in the network must first be grouped into fields of innovation. An innovation field contains a subset of the collected semantic network and comprises several terms or topics of the product environment. It is assumed that topics that are frequently mentioned together are interdependent. The interdependency of topics hints at their interrelatedness, therefore a clustering algorithm groups them in a common field of innovation.

In order to enable a targeted selection of fields of innovation by the company, they must finally be evaluated. The two suitable dimensions for evaluation are the external relevance of the field of innovation and the company's competence to carry out innovation projects addressing that specific field. The external relevance can be evaluated by using the potential for differentiation, expected long-term stability and strategic importance. The evaluation of competence must take into account the skills of the available employees, the available resources and, if applicable, previous experience in the field. In addition, the emergence of the topics in the external influence dimensions can be used as a decision support in order to incorporate specific external influences into the product innovation. Thereby, companies are enabled to address specific external influence dimensions such as customers or competition to innovate their product accordingly.

4 Conclusion

The effect of external influences on product development requires that companies observe the environment of their products holistically in order to identify necessary as well as promising possibilities for innovating the product. In particular, methods from the fields of environmental scanning, text mining and semantic networks offer the possibility to systematically evaluate even large amounts of data and to make them interpretable. However, current methods do not take advantage of this potential yet. The presented framework closes this gap by combining methods from the three areas and complementing each other to enable the identification of potential fields of innovation for existing products. The framework is a procedural model that presents five key steps necessary for the combined application of the methodology. Future research has to elaborate which specific influence dimensions from the product environment have to be taken into account for the identification of innovations. Further, the use of text mining can be improved by the development and application of topic models that specifically identify new topics that were not covered by previous documents to improve the indication quality.

Acknowledgements Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2023 Internet of Production—390621612.

References

1. Verein Deutscher Ingenieure: Entwicklung technischer Produkte und Systeme. Gestaltung individueller Produktentwicklungsprozesse, 1993rd edn. 03.100.40(VDI 2221 Blatt 2) (2019)
2. Chang, W., Taylor, S.A.: The effectiveness of customer participation in new product development: a meta-analysis. J. Mark. (2016). https://doi.org/10.1509/jm.14.0057
3. Riesener, M., Dölle, C., Becker, A., Schuh, G.: Modeling of usage hypotheses for the identification of innovation potential. In: 9th International Conference on Mass Customization and Personalization, pp. 191–198 (2020)
4. Andersen, N., Boersch, A., Blohmke, J.: Innovation in Europe. A Deloitte survey on European companies and how digital technologies can strategically enhance innovation. Deloitte Insights. https://www2.deloitte.com/de/de/pages/innovation/contents/innovationsreport-2019.html (2019)
5. Peter, M.K., Jarratt, D.G.: The practice of foresight in long-term planning. Technol. Forecast. Soc. Chang. (2015). https://doi.org/10.1016/j.techfore.2013.12.004
6. Enkel, E., Horváth, A.: Mit cross-industry-innovation zu radikalen Neuerungen. In: Ili, S., Albers, A. (eds.) Open Innovation umsetzen. Prozesse, Methoden, Systeme, Kultur, 1st edn., pp. 293–314. Symposion-Publ, Düsseldorf (2010)
7. Ur-Rahman, N., Harding, J.A.: Textual data mining for industrial knowledge management and text classification: a business oriented approach. Expert Syst. Appl. (2012). https://doi.org/10.1016/j.eswa.2011.09.124
8. Kölbl, L., Mühlroth, C., Wiser, F., Grottke, M., Durst, C.: Big Data im Innovationsmanagement. Wie Machine Learning die Suche nach Trends und Technologien revolutioniert. HMD (2019). https://doi.org/10.1365/s40702-019-00528-3
9. Robinson, C.V., Simmons, J.E.L.: Organising environmental scanning. Exploring information source, mode and the impact of firm size. Long Range Plann. (2018). https://doi.org/10.1016/j.lrp.2017.10.004
10. Theobald, E.: Marketing intelligence studie 2018 in deutschen B2B-Unternehmen. https://www.researchgate.net/publication/332116819_Marketing_Intelligence_Studie_2018_in_deutschen_B2B-Unternehmen (2019)
11. Schuh, G., Dölle, C.: Sustainable Innovation. Nachhaltig Werte schaffen, 2nd edn. Springer Vieweg, Berlin, Heidelberg (2021)
12. Hope, T., Chan, J., Kittur, A., Shahaf, D.: Accelerating innovation through analogy mining. In: Matwin, S., Yu, S., Farooq, F. (eds.) Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '17: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax NS Canada, 13.08.2017–17.08.2017, pp. 235–243. ACM, New York, NY, USA (2017). https://doi.org/10.1145/3097983.3098038
13. Reichenberger, K.: Kompendium semantische Netze. Springer, Berlin Heidelberg, Berlin, Heidelberg (2010)
14. Albright, K.S.: Environmental scanning. Radar for success. Inf. Manage. J. 38–44 (2004)
15. Miles, I., Saritas, O.: The depth of the horizon. Searching, scanning and widening horizons. Foresight (2012). https://doi.org/10.1108/14636681211284953
16. Sutherland, W.J., Bardsley, S., Bennun, L., Clout, M., Côté, I.M., Depledge, M.H., Dicks, L.V., Dobson, A.P., Fellman, L., Fleishman, E., Gibbons, D.W., Impey, A.J., Lawton, J.H., Lickorish, F., Lindenmayer, D.B., Lovejoy, T.E., Nally, R.M., Madgwick, J., Peck, L.S., Pretty, J., Prior, S.V., Redford, K.H., Scharlemann, J.P.W., Spalding, M., Watkinson, A.R.: Horizon scan of global conservation issues for 2011. Trends Ecol. Evol. (2011). https://doi.org/10.1016/j.tree.2010.11.002
17. Heinonen, S., Hiltunen, E.: Creative foresight space and the futures window: using visual weak signals to enhance anticipation and innovation. Futures (2012). https://doi.org/10.1016/j.futures.2011.10.007
18. Mendonça, S., Cardoso, G., Caraça, J.: The strategic strength of weak signal analysis. Futures (2012). https://doi.org/10.1016/j.futures.2011.10.004

19. Yoon, J.: Detecting weak signals for long-term business opportunities using text mining of Web news. Expert Syst. Appl. (2012). https://doi.org/10.1016/j.eswa.2012.04.059
20. Ernstsen, S.K., Thuesen, C., Larsen, L.R., Maier, A.: Identifying disruptive technologies. Horizon scanning in the early stages of design. In: Proceedings of the DESIGN 2018 15th International Design Conference. 15th International Design Conference, May, 21–24, 2018, pp. 1833–1844. Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia; The Design Society, Glasgow, UK (2018). https://doi.org/10.21278/idc.2018.0105
21. Schuh, G., Reinhart, G., Prote, J.-P., Sauermann, F., Horsthofer, J., Oppolzer, F., Knoll, D.: Data mining definitions and applications for the management of production complexity. Proc. CIRP (2019). https://doi.org/10.1016/j.procir.2019.03.217
22. Hippner, H., Rentzmann, R.: Text mining. Informatik Spektrum (2006). https://doi.org/10.1007/s00287-006-0091-y
23. Jo, T.: Text Mining. Concepts, Implementation, and Big Data Challenge. Studies in Big Data, vol. 45. Springer International Publishing, Cham (2019)
24. Feldman, R., Sanger, J.: The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, Cambridge, New York (2007)
25. Aggarwal, C.C., Zhai, C.: Mining Text Data, 2012th edn. Springer, US, Boston, MA (2012)
26. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. (2003). https://doi.org/10.1162/jmlr.2003.3.4-5.993
27. Griol-Barres, I., Milla, S., Cebrián, A., Fan, H., Millet, J.: Detecting weak signals of the future: a system implementation based on text mining and natural language processing. Sustainability (2020). https://doi.org/10.3390/su12197848
28. Quillian, M.R.: Semantic Memory. Defense Technical Information Center, Fort Belvoir, VA (1966)
29. Shastri, L.: Semantic Networks: An Evidential Formalization and its Connectionist Realization. Morgan Kaufman Publishers Inc., Los Altos, CA (1987)
30. Sowa, J.F.: Semantic networks. In: Shapiro, S.C. (ed.) Encyclopedia of Artificial Intelligence, 2nd edn. (1992)
31. Brachman, R.J.: On the epistemological status. In: Findler, N.V. (ed.) Associative Networks. Representation and Use of Knowledge by Computers, pp. 3–50. Elsevier Science, Burlington (1979)
32. Giarratano, J.C., Riley, G.: Expert Systems. Principles and Programming, 3rd edn. PWS, Boston (1998)
33. Russell, S.J., Norvig, P.: Artificial Intelligence. A Modern Approach. Always Learning. Pearson, Boston, Columbus, Indianapolis (2016)
34. Ahmed, S., Kim, S., Wallace, K.M.: A methodology for creating ontologies for engineering design. J. Comput. Inf. Sci. Eng. (2007). https://doi.org/10.1115/1.2720879
35. Shi, F., Chen, L., Han, J., Childs, P.: A data-driven text mining and semantic network analysis for design information retrieval. J. Mech. Des. (2017). https://doi.org/10.1115/1.4037649
36. Vijayarani, S., Ilamathi, M.J., Nithya, M.: Preprocessing techniques for text mining—an overview. Int. J. Comput. Sci. Commun. Netw. **5**, 7–16 (2015)
37. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python. Safari Books Online. O'Reilly Media Inc., Sebastopol (2009)

A Collaborative and Interactive Surface Concept for Early Stages of New Product Development—A Multi-stage Expert Study

Verena Lisa Kaschub, Reto Wechner, Katharina Dieterich, Ralf Lossack, and Matthias Bues

Abstract To come up with new ideas and innovations, product developers are faced with the challenge of finding a suitable way through a jungle of tools and methods, e.g., creativity tools, CAD software and project management applications. The goal of this paper is to develop a user interface concept for a supportive tool that integrates a variety of tools on an interactive surface to enable horizontal collaboration and interaction during creativity sessions in product development. Therefore, the user interface should not just support collaboration but also incorporate guidance to support product developers to find a suitable path through this jungle of tools and methods. To elicit requirements for such a supportive tool, an explorative multi-stage study design was chosen. In the first part expert workshops were conducted to create a design space. This space was further explored in a second phase, to derive relevant requirements. In the last phase, first visions of the supportive tool and possible design ideas were developed. The results of this study form a foundation to build a supportive tool for product developers that could pave the way to a new form of digital collaboration, provides a more holistic understanding of the development process and its activities and furthermore enhances the acceptance of creativity methods.

V. L. Kaschub (✉) · K. Dieterich
GSaME—Graduate School of Excellence advanced Manufacturing Engineering, University of Stuttgart, Stuttgart, Germany
e-mail: verena.kaschub@gsame.uni-stuttgart.de

K. Dieterich
e-mail: katharina.dieterich@gsame.uni-stuttgart.de

R. Wechner
University of Stuttgart IAT, Stuttgart, Germany
e-mail: reto.wechner@iat.uni-stuttgart.de

R. Lossack
Siemens Industry Software GmbH, München, Germany
e-mail: ralf-stefan.lossack@siemens.com

M. Bues
Fraunhofer IAO, Stuttgart, Germany
e-mail: matthias.bues@iao.fraunhofer.de

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_69

Keywords Interactive surface concept · Creativity methods · Supportive tool · Tool integration · Product development

1 Introduction

To come up with new ideas and innovations, product developers are faced with the challenge of finding a suitable way through a jungle of tools and methods, e.g., creativity tools, CAD software and project management applications. Every tool is developed to support them during ideation and collaboration, but it only covers a part of the process. Therefore, product developers are faced with dealing with multiple tools simultaneously. The usable space for software tools on common work environments is usually limited by displays and projections on personal desks or walls in conference rooms. This leads to application switching and/or splitting up the already limited screen space. Furthermore, due to the vertical arrangement of these screen spaces and limited interaction usually performed by one presenter, collaboration takes place in a hierarchical manner and impacts the interaction of a team [1].

Moreover, during the early stages of product development product developers have a variety of different methods at their disposal and new methods are continuously being developed and established [2]. This leads to a "method jungle" in which it is difficult for product developers to find the most suitable way to solve a problem [3]. Despite the added value of methods in product development (structuring process, enabling creativity, process documentation, preventing errors) [4–6], studies show that only low usage is observed in the development process [3, 7]. Reasons for low usage include limited experience and lack of knowledge of the individual methods. Therefore, it is essential to not only provide a suitable interface and surface solution to enable mutual interactive space to collaborate on but also to incorporate guidance to support product developers in this jungle of methods and tools to find a suitable path.

To dissolve these restrictions, an interactive surface that enables horizontal collaboration and interaction during creativity sessions in product development, integrating a variety of tools (e.g., whiteboard, CAD) to collaborate with simultaneously, is proposed. In this paper requirements are elicited to visualize a supportive tool and present a first concept of how product developers could use and interact with such a supportive tool.

In the first part of the paper, the method with the explorative study design is illustrated that spans up a design space and visualizes first ideas for a concept of the supportive tool. In the second part of the paper the results of these studies are presented. The paper concludes with a discussion of the results and proposes future research.

2 Method

2.1 Stage 1: Span Design Space

To elicit requirements for a supportive tool, a multi-stage explorative study design was chosen. In the first part of the study, digital expert workshops were used to specify a design space. This space is defined by a variety of dimensions, for example size, color or tactile quality of a product. Dimensions can also be more complex, such as interaction and connectivity. These dimensions span the design space to provide a framework for further exploration (see Fig. 1) [8].

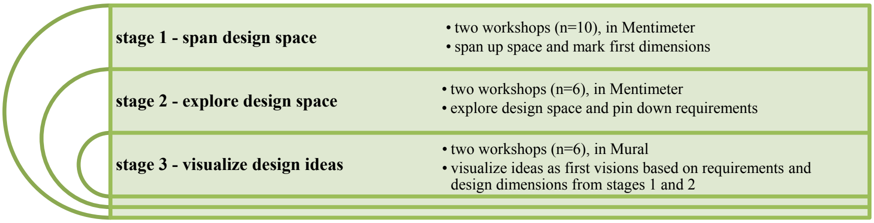


Fig. 1 Study design of the three stages of expert workshops

During the first stage of the study, two digital expert workshops were conducted (approximate duration of 60 min) using the digital interactive presentation tool Mentimeter (due to Corona restrictions only digital format was possible) [9]. To evaluate the format of the workshop, a pilot study was executed and the order as well as the formulation of the questions was changed to derive a better comprehensibility. Afterwards two workshops with in total ten experts (3 female, 7 male) with a mean age of 41.51 years (SD = 10.52) were held. The area of expertise varied from mechanical engineering, innovation management to computer science. All participants had a master's degree, three of them a PhD and one was a professor.

During the workshops the experts had the opportunity to state their visions and ideas and to discuss them in direct exchange within the group. The questions addressed different dimensions of the design space: the UI and its communicational aspects, the functionality of a supportive tool; the physical basis/hardware and the interoperability to other tools.

Afterwards two researchers analyzed the results separately and condensed them to the main dimensions (hardware and communication/UI). Then, the analyses were compared and discussed to generate a devised vision of the design space dimensions.

2.2 Stage 2: Explore Design Space

In the second phase of the study, this design space and its dimensions were explored further to elicit specific requirements for the different dimensions. The aim was not only to derive relevant requirements but furthermore form a grasp of a first vision for a supportive tool concept. In total two short workshops were executed (duration of approximately 20 min) to ensure that participants did not get bored and exhausted by answering questions in a digital setting (this feedback was given during the first sessions). These workshops also took place digitally with Mentimeter. Six participants took part in the workshops (5 male, 1 female) with a mean age of 38.83 years (SD = 12.78). The areas of expertise were computer science and UI/UX with one expert holding a PhD and one being Professor.

To get the participants familiar with the topic and make it easier to empathize with the needs of a possible end-user, a fictional persona was used. This persona was called Charlotte, a product developer that tried to find a solution with her team members on what the interior of a future car should look like. To challenge the participants and set them free from boundaries of today's technology, the problem setting of Charlotte took place in 2030. The short introduction with the storyline was finished with a picture of Charlotte and her team members 'standing in front of a method-jungle' and not being able to find the best way to solve the problem. The question that introduced the active part of the workshops was how a future supportive tabletop tool (tabletop hardware was chosen due to results from first stage) can assist Charlotte and her team members. During the second session the participants were reminded by the persona to not lose context. Then the active part of the workshops began.

In total 18 questions were asked during the workshop (11 in the first session, 7 in the second session). The questions were directed towards interoperability and integration of tools, UI design and integration of procedural process guidance with methods. Closed-ended question were applied to pin down requirements (e.g., how many participants should work in a creativity session when using the tabletop?). Moreover, to ensure an exchange between experts and provide the option to elaborate on ideas, open-ended questions were chosen when the requirement was still too vague as well as to get a broader view on the given ideas. An example of an open-ended question was what kind of information should be presented to users at the startup of the tool. After the two workshop sessions had been held, two researchers analyzed the results and set up a mind map that structured the design space with its dimensions and implementation ideas.

2.3 Stage 3: Visualizing Design Ideas

The third and last phase of the multi-stage expert study was focused on forming a first visions of a supportive tool and visualize possible design ideas. The aim was to develop initial visions further and derive a concept, based on the design dimensions and their requirements. The first two stages were conducted in Mentimeter due its structured presentation capabilities. For stage three, more degrees of freedom were needed to design options for a concept and not be held back by the limits of a tool like Mentimeter, therefore the online whiteboard-tool MURAL was used [10]. This made it possible to address specific questions about requirements as well as to collaboratively prototype concept ideas on a shared canvas. The questions in the workshop referred to specific dimensions of the design space (e.g., which tools should interact with each other). Also, during the third stage, the workshop was divided into two sessions (approximately 30 min) to keep up the attention of the participants.

During the first workshop session the participants from stage two were introduced to the results by presenting the mind map which was derived from the results of the first two stages. Afterwards, two design constructs were explained that could aid a future concept idea. The concept of nudging and the concept of gamification [11, 12] were introduced via short video clips, explaining the construct as well as showing examples of their implementation. The participants were asked to think about a possible integration of these constructs, to enlarge the acceptance as well as user experience of the concepts. The last step of this session was an icebreaker to the tool MURAL. Due to the higher degree of freedom of the tool (compared to Mentimeter) the participants did a competition by completing small tasks in MURAL that got them familiar with the functionalities (e.g., drawing, pasting pictures, writing).

The second session then solely focused on developing a vision of a tabletop concept. After having been equipped with all material (mind map, videos) the participants were divided into two groups (three participants each) and began to draw up their vision of a concept (see Fig. 2). Two researchers accompanied the session, guided if needed and supported the group with the functionalities of the tool MURAL.

To analyze the last stage, the new requirements (derived from the prototyped concepts) were added to the mind map. The specifications of the different stages were then evaluated and discussed to draw up a synergized and concise vision of a

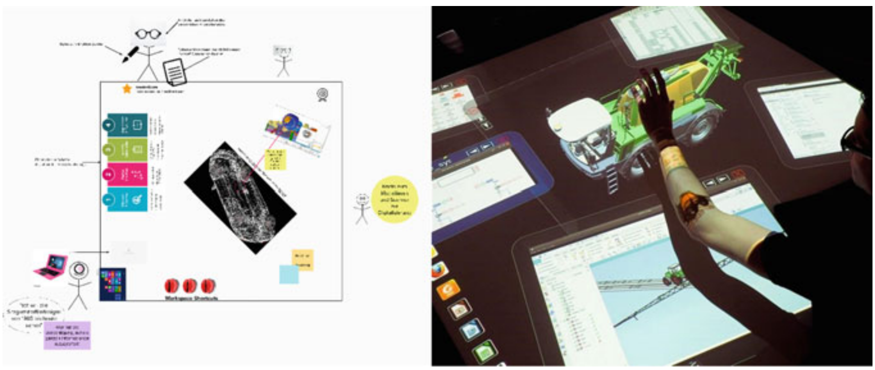


Fig. 2 Left: a digital visualized concept of third stage workshop; right: first synergized prototype of supportive tool

supportive tool within the design space and its dimensions. This synergized vision was prototyped in a first tabletop concept (see Fig. 2).

3 Results

3.1 Stage 1: Design Space Dimensions

The first stage of the study focused on setting up the dimensions of the design space. During the workshops four main dimensions were identified (Hardware, Automation, Interaction and Functionalities). One on these main dimensions was the physical appearance and its hardware. The results suggested digital as well as analog ideas. The tool was visualized as a kind of analog 'method cupboard' equipped with all material (e.g., sticky notes, play dough…) needed to achieve a successful creativity session as well as an expert guiding these sessions. The same vision was also present when the digital format was construed. Here the focus was on the adaptiveness of a digital avatar that accompanies and supports the process which could be directed towards an AI. The hardware base of a digital supportive tool should not only focus on one end-device but should be accessible via multiple devices, to cover collaborative settings but also decentralized collaboration. The tool was preferably viewed via a web/smartphone-application for quick access and enabling users to engage from a distance and on a tabletop hardware for collaborative settings (tabletop that goes 'beyond the typical tablet setting').

To be able to interact with the hardware, the dimension 'interaction' was elaborated on. The interaction between the tool and the users should be multi-channeled (auditive, visual and haptic). Auditive and visual elements can be combined to visually support a dialog. Visual elements can be as minimal as icons and transform towards complex graphics and holograms. To receive feedback or enter information to the tool, haptic elements can be applied (ultrasound and touch). A given example for a possible interaction was the quiz game jeopardy with IBM Watson. To be able to communicate on a very specific level (with users that are directly involved in the development) and on a more generic level (e.g., status meeting with management) different levels of representation and communication should be provided by the tool.

The third dimension illustrated the functionalities of the tool. These functionalities were discussed via the output options of the tool. These options were directed towards supporting product developers with creativity, collaboration methods and content-related issues. The given output should be adaptive and match, for example, team members with suitable experts or find the most fitting creativity method to the situation and suggest matching tools. The content-related output is focused on providing solution patterns and associative background information with best-practice solutions. This output should be unobtrusive and just be presented when required compared to the methodological support which should guide product developers more actively through the process. To provide output, the aspect data was added to the functionality dimension. The data should hold information on who is involved and the associated user profiles as well as constraints (e.g., goals, time limits). To be able to actively support creativity methods an overview of available methods and their descriptions should be integrated. Product data and ontologies can aid to provide content-related support. The needed data should be acquired manually as well as automated. To acquire data, interfaces to other tools are essential. Important tools that were mentioned can be categorized as project management tools or design and creativity tools.

3.2 Stage 2: Dimensions and Requirements

The second stage elaborated on the dimensions of stage two and gave design suggestions. Moreover, the dimensions were further marked down and transformed to the three main dimensions: 'general design', 'tools' and 'methods' (which were broken down and clustered). The 'general design' dimension was clustered into user interface design, spatial dimension and general implications. The interface should be structured and minimalistic to be unobtrusive (e.g., more graphical than textual information). An overview of the process (e.g., progress bar, milestones) and goals should be illustrated. To ease the interaction, tangibles should be integrated such as clay, play dough or sticky notes.

To prevent the interaction and collaboration of the future tabletop concept exceeding its capabilities, experts decided that not more than five persons should collaborate on a tabletop (but at least two persons). During a creativity session a mixture of, only the moderator being able to interact and the whole team can interact with the tabletop, should be implemented to further structure the sessions. These functions should be communicated clearly at the beginning of a session/or during first usage to prevent false expectations. The interface should be structured into a common area which is centered and individual areas that are close to the users. These areas can either function within the tool (e.g., stylus, touch or keyboard and mouse on the tabletop) or be integrated in external hardware (e.g., laptop or tablet).

The second dimension 'tools' is clustered in three areas. The first being the functionalities. Project management tools should be integrated with support on collaboration setups (choosing suitable team members/experts) and structure the collaboration with tools on task management. The area of implementation describes that the functionalities of the tool (e.g., creating digital sticky notes or scribbles) should either be implemented within the tool itself or integrate already established software solutions. The cluster of interoperability states that the described functionalities should be interoperable with each other and enable the integration of external data.

Due to the relevance of methods in a creativity process, it became a separate dimension. Here the focus was put on explanation of the methods and in which structure they should be provided (as a whole in the beginning and then step by step when needed). The explanation should be executed over different channels (audio and visual) and should be oriented on current needs (to not overwhelm users with input). To be able to comprehend the decision of which method is provided by the supportive tool, advantages and disadvantages of the method should be presented as well as the goal of the method.

Hence, creativity methods should be selected by the supportive tool and then be provided to product developers, data for a possible automated selection was suggested within the method dimension. The tool should learn from previous sessions and user behavior to conduct a fitting strategy of method selection for future sessions. Therefore, the tool needs to be equipped with functionalities that are able to 'listen and watch' the process. Not only the user behavior should be taken into account, but furthermore data from the project itself (such as deadlines and time limitations) can guide the method selection. Lastly, the method dimension should focus on the documentation of content that emerges from working with the methods. This should be accessible within the tool but also from external sources (if permitted).

3.3 Stage 3: Digital Tabletop Concepts and Requirements

The last stage integrated previous requirements and their dimensions into two possible concept ideas of tabletops. These concepts went a step further and illustrated details and other important design elements which were not yet listed. As an example, one concept is shown in Fig. 2. In the following sections, requirements of both concepts are presented.

When actively analyzing the actual process of the supportive tool, the design issue of switching between workspaces e.g., project-management, designing and creativity environments became prominent. A workspace switcher was therefore integrated that enabled the shift between workspaces. Digital as well as analogous elements were mentioned such as an 'app-viewer' and a buzzer.

Furthermore, the separation between personal spaces and common areas was drawn up in more detail. In order to be able to keep the personal space truly 'personal', the visibility should only be available via AR for the person using it, but should integrate the option to enable team-view. This personal space should contain tools and functionalities that are chosen based on users' current needs (e.g., to-do list vs. scribble space). The personal and common space should also differ in the permissions of interactions e.g., users can access content of other personal spaces if their owner allows it. When talking about the common area, an important aspect seems to be to have the ability to connect related fragments of different tools (e.g., 2D CAD element with 3D CAD view and connected notes/sticky notes). This can be realized with a connecting line or a color coordination.

Another dimension that was further specified was the integration of non-digital elements such as sticky notes or play dough. Here a scanner located above the tabletop concept should be integrated to document these elements to save different development states and be able to share these elements during decentralized collaboration.

When integrating tools, it is important to show an overview on the current 'status' such as unfinished and future steps (e.g., with creativity methods). Not only the overview is relevant, but also the permission of elements within tools. Not every information should be accessible for every session/person.

All these design requirements that were mentioned during the workshop put additional emphasis on the unobtrusive configuration of a concept. Design ideas to apply this reserved view was to enable different degrees of support regarding the experience and skill level of the user as well as the option to only trigger additional support when asked for.

A first vision of the concept is visualized in Fig. 2 (right figure). It integrates CAD tools with common desktop applications (e.g., to-do list and creativity tools/methods) in a collaborative tabletop environment. This was implemented by using the VD1 desktop environment [13].

4 Discussion, Future Research and Conclusion

The goal of this paper was to develop a user interface concept that integrates a variety of tools on an interactive surface to enable horizontal collaboration and interaction during creativity sessions in new product development. The multi-stage expert study provided a first vision of a tabletop concept. Furthermore, it elicited requirements that not only determine a design space but provide inspiration for future research. The different dimensions (e.g., design, method, tool, hardware…) and the associated design ideas and requirements should be elaborated on. Design ideas and requirements like the distinction of private and common areas and the app switcher functionality could be implemented by extending the VD1 desktop environment. To fulfill the requirement of scanning sticky notes and other haptic/non digital elements already existing applications (e.g., Post-it® App, sticky notes capturing by Miro, Trnio App, Qlone App…) could be evaluated and potentially integrated into the tool. Furthermore, some kind of digital assistant should be implemented to guide product developers in their creativity process (e.g., method selection, method explanation…). These requirements form the foundation to build a supportive tool for product developers that tackle different shortcomings of conventional settings (e.g., team interaction…). The supportive tool could provide a more holistic understanding of the development process and its activities by integrating a variety of tools as well as providing guidance through the "method jungle". This could pave the way to a new form of digital collaboration and furthermore enhance the acceptance of creativity methods.

Due to the setting of the study the results must be viewed with their limitations. The applied tools to execute the workshop restricted the experts in their degree of freedom due to their inexperience. It might be that different concepts would have been developed/evolved if participants had the opportunity to use material that they are familiar with (e.g., paper-prototyping).

Future research should integrate users/product developers directly into the workshops to be able to evaluate the dimensions and requirements with them and identify the most suitable solution. The ideas and requirements can only be fully understood when users can test them in real life settings. The functionalities of the 3D integration (see Sect. 3.3) might feel different when experienced (evaluation via UX/Usability scales). Although this integration in VD1 is possible, future research should focus on the challenges integrating interactive functionalities into existing closed source 3D software (e.g., Siemens NX). Therefore, future research should incorporate these aspects and further prototype first solutions such as visualized in Fig. 2. In conclusion it can be said that this study proposed a first vision of supportive tools that enables future research to elaborate and evaluate on.

References

1. Al-Megren, S., Kharrufa, A., Hook, J., Holden, A., Sutton, S., Olivier, P.: Comparing fatigue when using large horizontal and vertical multi-touch interaction displays. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 156–164. Springer Verlag (2015)
2. Becerril, L., Guertler, M., Longa, E.: Developing design methods—a conceptual requirement framework. Proc. Des. Soc. Int. Conf. Eng. Des. **1**, 1463–1472 (2019). https://doi.org/10.1017/dsi.2019.152
3. Reiß, N.: Ansätze zur Steigerung der Methodenakzeptanz in agilen Prozessen der PGE—Produktgenerationsentwicklung (2018)
4. Lindemann, U.: Methodische Entwicklung technischer Produkte. Springer, Berlin Heidelberg (2009)
5. Pahl, G.: Psychologische und pädagogische Fragen beim methodischen Konstruieren: Ergebnisse des Ladenburger Diskurses von Mai 1992 bis Oktober 1993. Verlag TÜV Rheinland (1994)
6. Ehrlenspiel, K., Meerkamm, H.: Integrierte Produktentwicklung. In: Integrierte Produktentwicklung, pp. I–XXV. Carl Hanser Verlag GmbH & Co. KG (2013)
7. Yeh, T.-M., Pai, F.-Y., Yang, C.-C.: Performance improvement in new product development with effective tools and techniques adoption for high-tech industries. Qual. Quant. **44**, 131 (2010)
8. Flemisch, F., Semling, C., Heesen, M., Meier, S., Baltzer, M., Krasni, A., Schieben, A.: Towards a balanced human systems integration beyond time and space: exploroscopes for a structured exploration of human-machine design spaces (2013)
9. Mentimeter. https://www.mentimeter.com/ (2021)
10. MURAL. https://www.mural.co/ (2021)
11. Thaler, R., Sunstein, C.: Nudge: Improving Decisions About Health, Wealth, and Happiness (2009)
12. Sailer, M., Hense, J., Mandl, J., Klevers, M.: Psychological perspectives on motivation through gamification. Interact. Des. Archit. J. 28–37 (2014)
13. Bues, M., Wingert, B., Riedel, O.: VD1: a technical approach to a hybrid 2D and 3D desktop environment. 2018 IEEE 11th Workshop on Software Engineering and Architectures for Real-Time Interactive System SEARIS (2018). https://doi.org/10.1109/SEARIS44442.2018.9180231

Interorganizational New Product Development: A Future Vision of Project Team Support on an Organizational, Relational, and Content-Related Collaboration Level

Katharina Dieterich, Verena Lisa Kaschub, and Peter Ohlhausen

Abstract Because of a high product and technology complexity, companies involve external partners in their research and development (R&D) processes. Interorganizational projects result, which represent temporary organizations. In these projects heterogenous organizations work closely together. Since project work is always teamwork, these projects face due to their characteristic's major challenges on an organizational, relational, and content-related collaboration level. Thus, this paper raises the following research question: "How can a project team be supported on an organizational, relational, and content-related level in an interorganizational new product development setting?" To answer this research question, an explorative expert study was set up with two digital workshops using the interactive presentation tool Mentimeter. The results show that a cooperative innovation culture could support project teams on an organizational and relational level in the future in minimizing predominant problems. Moreover, it supports project teams for example in a functional communication. Furthermore, 18 values of a cooperative innovation culture result which are for example openness and transparency, risk and failure tolerance or respect. On a content-related level the results show that an adaptable tool which promotes creativity and collaboration method as well as content-related input support could be beneficial for problem-solving in an interorganizational new product development setting in the future. Because the tool can guide product developers through the process with suitable creativity and collaboration methods, can give content-related input and can enable interactive interchange on a table-top. Future research

K. Dieterich (✉) · V. L. Kaschub
GSaME—Graduate School of Excellence advanced Manufacturing Engineering, University of Stuttgart, Stuttgart, Germany
e-mail: katharina.dieterich@gsame.uni-stuttgart.de

V. L. Kaschub
e-mail: verena.kaschub@gsame.uni-stuttgart.de

P. Ohlhausen
Reutlingen University and Fraunhofer Institute for Industrial Engineering IAO, Reutlingen and Stuttgart, Germany
e-mail: peter.ohlhausen@iao.fraunhofer.de

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_70

could mainly focus on the connection of the cooperative innovation culture and the tool since these potentially influence each other.

Keywords Interorganizational new product development · Project team support · Collaboration levels

1 Introduction

Because of a high product and technology complexity, companies involve external partners in their research and development [1, 2]. For solving complex problems firms seem to prefer nonequity partnerships with organizations which contribute complementary knowledge [3]. Nonequity partnerships are contractual agreements without equity ownerships and thus have high opportunism risks [3, 4]. An interorganizational, temporary organization results in which heterogenous organizations work closely together [5, 6]. Project work is always teamwork [7]. Hence, project teams face due to the projects' characteristics major challenges in new product development. According to teamwork takes place on the three collaboration levels 'organization', 'relation', and 'content' [7]. The levels 'organization' and 'relation' focus on working on a social system, which is here an interorganizational project. The organizational level comprises the project organization, the connection to the parent organization as well as structures, processes, and methodological guidelines [7]. To the relational level shaping the relationship, trust, personal communication, project culture as well as dealing with conflicts can be assigned. The content-related level focuses on working within a social system and comprises productive work, scope, time, requirements, decisions, concepts and cost [7].

In an interorganizational new product development setting, these collaboration levels become relevant for problem solving in teams [8]. Since there are interactions between these three collaboration levels [7], all three levels must be considered for supporting interorganizational project teams in their collaboration. Thus, this paper raises the following research question: "How can a project team be supported on an organizational, relational, and content-related level in an interorganizational new product development setting?".

The structure of the paper is as follows. Section 2 describes the method on how the research was conducted. Afterwards, in Sect. 3, the results are presented, following the collaboration levels of [7]. In Sect. 4 the results are discussed and then in Sect. 5 a conclusion, limitations as well as implications for future research are given.

2 Method

To take a closer look on how project teams can be supported on an organizational, relational, and content-related level in an interorganizational new product development setting, an explorative expert study was set up. A workshop design was chosen in order to reflect different opinions from different disciplines on this research topic. The goal of the workshop was to define current challenges and problems as well as determine possible ways of support on each collaboration level. Due to the current situation the expert workshops were executed digitally. The interactive presentation tool Mentimeter [9] was used to engage with the participants from distance. Before the two workshop sessions were executed a pilot-workshop was held with three participants to evaluate the comprehensibility of the questions as well as to account for the duration of the workshops. With minor changes two workshops then took place with a duration of approximately one hour. In total 10 experts took part in the study (5 experts per workshop, 7 male and 3 female), with a mean age of 41.5 years (SD = 10.52). The area of expertise varied from computer science and mechanical engineering to innovation- and project-management. All participants had a master's degree, three of them a PhD and one was a professor. The participants came from research as well as industry and are aware of the particularities of an interorganizational new product development setting. Due to their background, they can deal with the settings' complexity and bring up different perspectives on project team support.

After signing a document on data protection and a short explanation of the tool Mentimeter, the experts were then introduced to the workshop topics. The workshops were divided into two parts. Both parts consisted of open as well as closed-ended questions, so participants had the choice to elaborate on their previous given answers. Questions were directed towards the three collaboration levels of: content-related, relational and organizational level [7].

The first part dealt with the content-related level. The questions were focused on a vision of a 'tool' that could support product developers on a content-related level in the future. To assure that participants got a better grasp of the vague topic, the part was introduced by a short framework. The questions started with construing a generic view of a future tool (what should be supported?) and elaborated on the topic on a more detailed level (e.g., how does a possible output of the tool look like? How intensive should the support be on different areas?). In general, the questions of this first part dealt with general requirements that are essential when supporting new product development processes on a content-related level.

In contrast to the first part, the second part of the workshop covered the other two levels of the organizational and relational level [7]. The questions were focused on a vision of a cooperative innovation culture (see [8]), that could support interorganizational project members on an organizational and relational level in the future. To assure that the participants were on the same level for the following questions, the terms innovation network, cooperative innovation culture and values were explained (for definitions of network, culture and values see [8]). Afterwards, four questions on network characteristics of interorganizational projects on firm level were asked, as these have an impact on the project culture development (see [8]). These questions focused on the dimensions from who elaborated a morphological-typological systematization framework for firm networks [10]. This framework comprises six dimensions [10]. However, the experts were only asked about the four dimensions transaction link, strategy focus, order principle and degree of formalization as the other two dimensions were already answered by the project characteristics (see [10]). The four remaining dimensions have dichotomous items which were mapped on a 5-point-Likert-Scale. The scales were described verbally to the experts before they answered the questions using mostly the explanations of [10]. Only for order principle, the explanations of were used, since he focuses on the design of the network organization and its business processes [11].

Afterwards, the focus was laid on problems that can occur when working in an interorganizational project and on how a cooperative innovation culture could support here. Then, the experts were asked about the values of a cooperative innovation culture. To finish the workshop and conclude and reflect on questions beforehand, the importance of future support on the three levels was asked. Then, the participants had the opportunity to rate the format of the workshop as well as give feedback for future improvement.

During the workshops two researchers took notes on what was mentioned when discussions or questions to single points came up. After each workshop, the notes were compared so that the emerging results were objectified. Then, the data was analyzed and translated from German into English.

3 Results

The following results are structured according to in organizational and relational collaboration level and in content-related collaboration level [7]. Since the experts evaluated project team support on an organizational and relational collaboration level [mean value for each: 3.9/5] as more important than on a content-related collaboration level [mean value: 2.8/5] in an interorganizational new product development setting, these levels are presented firstly.

3.1 Organizational and Relational Collaboration Level

Networks on firm level in interorganizational projects seem to be interdependent [mean value: 3.1], rather problem-solving-oriented [mean value: 3.7], rather heterarchical [mean value: 3.9] and have a rather low degree of formalization [mean value: 3.6]. In an interorganizational new product development setting different problems can occur. Communication problems and dysfunctional communication appear. Project teams use for example different terminologies, definitions or technical language. Moreover, the protection of the intellectual property is an issue and project members are afraid of theft of ideas. In addition, a lack of trust or missing trust occurs in an interorganizational new product development setting. Other problems come from hidden agendas, non-transparence, e.g., in interests, competition and differences in goals of the project partners. Moreover, cultural differences, different values and different personalities can be problematic. Furthermore, different engagement and appreciation towards the project as well as hierarchies, different time horizons and spatial distances can be issues. Another problem comes with the project size; the bigger the harder it is. In addition, money, politics and a high fluctuation can be problematic. Also, different time zones and languages can be other problems.

A cooperative innovation culture can support project teams in several ways. The experts indicate that a cooperative innovation culture can build and enhance trust among the project members. They mentioned trust several times, therefore it seems to be important. Moreover, it supports project teams in setting mutual values, common customs and rules of conduct, so that project team members have a sense of belonging and a desire to cooperate with each other. In addition, a cooperative innovation culture supports project members in a mutual language and functional communication as well as in promoting team spirit and team building. Cultural differences are bridged, personal relationships are maintained and openness as well as respect is promoted. Furthermore, a cooperative innovation culture supports in conflict solving and sets other incentive systems than money. Moreover, it supports project members in freedom concerning time and resources as well as in sustainability. In addition, it can support project members by giving them a guideline on a methodological level on how to develop trust.

Besides, the experts named different values of a cooperative innovation culture. They indicate trust, openness, tolerance and resilience as values. Moreover, they named respect, reliability, empathy, fairness, curiosity and honesty. In addition, they stated that an error culture, an open communication, freedom, out-of-the-box and trying out are further values. The experts named trust, openness and respect as well as an error culture more than once. Hence, these values seem to be important. Moreover, a mutual goal, creating value and personal development are other values they named. In addition, give not only take and interest toward the topics of the project partners are further values of a cooperative innovation culture. Furthermore, some results were rather describing, e.g., concerning benefits of a mutual work or concerning values and their creation. For example, it was stated that values are created in a project team, a mutual value understanding is a basis of a team, each team member should try to comply with those and that a resulting construct of values is dynamical.

To sum up, on an organizational and relational level a cooperative innovation culture can support project teams in minimizing predominant problems. Its values play a major role which are for example respect, openness or tolerance.

3.2 Content-Related Collaboration Level

The vision of a tool that support new product development within a social system is elaborated on different aspects that are essential to set up a first concept vision. In general, a tool support for product developers should focus mainly on creativity and collaboration method guidance but also provide some content-related support. When thinking about main requirements that a tool should be equipped with, adaptability is a key point. Adaptability emerges with various design components. When interacting with the users, the tool should be adaptable and for example focus on the experience level of product developers to adapt the output accordingly. Furthermore, the hardware should also be adaptable. When the team works decentralized, the tool should work on hardware that is accessible within these situations (smartphone/laptop). Compared to a collaborative in-person setting, the tool should function on a table-top, that enables an interactive interchange.

The output of the tool that is then provided to product developers should focus on creativity and collaboration methods as well as provide content-related input to the users. Nevertheless, the intensity should vary within these output options. While content-related input should be minimal the support for creativity methods should take place in a more guided format. When visualizing the kind of output that a tool provides to product developers, different options should be allocated. The creativity method output could identify different solution method-paths and equip users with 'ready-to-use' instructions that guide them step-by step through the methods. This guidance should be intuitive and easy to follow. The collaborative method support should interact in a similar manner. Here the guidance is focused on figuring out and determining suitable project partners and experts. The tool should then give suggestions on how to set-up the interaction and guide users through the process. Compared to the creativity and collaboration method support, the content-related support provides no guidance path, but support product developers with a direct input of similar ideas and patents related to the current problem discussed. Moreover, it shows market data on competitors and could provide legal advice.

To sum up, on a content-related level, a tool should support with collaboration and creativity methods as well as in giving content-related input. Focus should lie on the adaptability to cater to all product developers.

4 Discussion of the Results

The results show mainly the importance of mutual values for interorganizational project team support on an organizational and relational collaboration level. This is also supported by who showed that values of a cooperative culture and an innovation culture are the basis for a cooperative innovation culture [8]. Hence, to further elaborate on how a project team can be supported on an organizational and relational collaboration level, the underlying values of a cooperative innovation culture are important. Values are desirable, abstract action guidelines [12]. Some values, the experts named do not fit with this common definition. That's why the mentioned values are mirrored against values from literature. Since there is to the best of our knowledge no evidence on a value basis of a cooperative innovation culture, these values are mirrored against values of a cooperative culture and values of an innovation culture. So far, there is less evidence on values of a cooperative culture. Following the adapted reasoned action approach for interorganizational projects (see [8]), values of a cooperative culture need to promote trust. Hence, trust is not a value, as the experts indicated here, but an intention for a cooperative behavior (see [8]). Butler 1991 and Schön 2020 show different factors promoting interpersonal trust [13, 14]. These factors fit with the definition of values given above and can therefore be seen as values for a cooperative culture. For innovation culture, there is much evidence on its values. In a qualitative meta-study of 110 research articles, identified twelve value dimensions of an innovation culture [15]. With mirroring the values named by the experts (without the describing results) against these values, 18 values of a cooperative innovation culture result (see right column in Table 1). These values consider the experts view and are valid within the framework of the common definition, given above. Hence, for supporting project teams with a cooperative innovation culture in the future it is essential for the project team to comply with these 18 values. Based on these, the definition of the valid norms in the project team and the design of the operational and organizational structure can then take place (see [8]).

5 Conclusion, Limitations and Future Research

This paper aimed at answering the following research question: "How can a project team be supported on an organizational, relational, and content-related level in an interorganizational new product development setting?" To answer this question an expert study was set up with two digital workshops, using the interactive presentation tool Mentimeter. The results give an overview of network characteristics on firm level in interorganizational projects. Moreover, the results show that in an interorganizational new product development setting project team support on an organizational and relational level is more important than on a content-related level. On an organizational and relational collaboration level the development of a cooperative innovation culture supports project teams in minimizing predominant problems which concern for example communication, intellectual property or cultural differences. A cooperative innovation culture supports project teams for example with a functional communication, openness or in establishing mutual values and rules of conduct. Moreover, 18 values of a cooperative innovation culture result which are for example risk and failure tolerance, openness and transparency or respect. On a content-related level, the results show that a tool is necessary which supports the project team with creativity and collaboration methods as well as with a content-related input in the future. The main requirement for this tool is adaptability for an interaction with the users, for the output as well as for the hardware.

Table 1 Values of a cooperative innovation culture

Values, named by the experts	Values of a cooperative culture*	Values of an innovation culture**	18 resulting values of a cooperative innovation culture
Tolerance	-	Risk and failure tolerance	Risk and failure tolerance Openness and transparency Respect Willingness to cooperate Newness and creativity Focus Resilience and flexibility Fairness Technology orientation Integrity Autonomy Future orientation Market orientation Learning orientation Consistency Promise fulfillment Appreciation Interest
Error culture	-	Risk and failure tolerance	
Tolerant error culture	-	Risk and failure tolerance	
Trying out	-	Risk and failure tolerance	
Openness	Openness, transparency	-	
Openness	Openness, transparency	-	
Open communication	Openness, transparency	Willingness to cooperate	
Give not only take	-	Willingness to cooperate	
Respectful interaction with each other	Respect	-	
Respect	Respect	-	
Respect	Respect	-	
Resilience	-	Flexibility *(partly)*	
Fairness	Fairness	-	
Curiosity	-	Newness and creativity, Technology orientation	
Out-of-the-box	-	Newness and creativity	
Honesty	Integrity	-	
Freedom	-	Autonomy	
Value creation	-	Focus, Future orientation, Market orientation	
Mutual goal	-	Focus	
Personal development	-	Learning orientation	
Reliability	Consistency, promise fulfillment	-	
Empathy	Appreciation	-	
Interest towards partners' topics	Interest	-	

*see Butler 1991 and Schön 2020; **see Rieger 2018

With these results this paper provides a future vision on project team support on three collaboration levels in an interorganizational new product development setting. However, there are a few limitations of the paper. The workshop took place digitally. That's why the study had a high degree of abstraction and thus the grasp of the problem definition was difficult for the participants. Moreover, the workshops were more general oriented. Thus, no concrete use case is illustrated that can elaborate on the single statements. It would be interesting for future research to compare whether there are different statements when they are focused on a concrete use case. Moreover, future research should focus especially on the connection of the support possibilities on the three collaboration levels. Since the three collaboration levels influence each other, this is as well the case for the individual support possibilities. Hence an interesting future research question could be: "How does a tool which supports product developers with creativity and collaboration methods as well as with content-related input influence a cooperative innovation culture in an interorganizational new product development setting?" To answer this future research question, there is a closer look on the concrete design of the tool on the content-related level necessary. Because, while there is a rather concrete vision of project team support on an organizational and relational level, for a concrete future vision on a content-related level it could be interesting for future research to examine how the tool functions in more detail. Besides, future research could evaluate the resulting 18 values of a cooperative innovation culture.

Acknowledgements We would like to thank all experts who participated in our workshops. Since some experts work within the ARENA2036 environment, we would also like to thank the ARENA2036 that we could conduct some of our research there.

References

1. Bullinger, H.-J., Warschat, J.: Innovationsmanagement in Netzwerken. In: Garcia Sanz, F.J., Semmler, K., Walther, J. (eds.) Die Automobilindustrie auf dem Weg zur globalen Netzwerkkompetenz. Effiziente und flexible Supply Chains erfolgreich gestalten, pp. 199–214. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg (2007)
2. Fraunhofer-Verbund Innovationsforschung: Wandel verstehen—Zukunft gestalten. Impulse für die Zukunft der Innovation. Stuttgart (2018)
3. Bagherzadeh, M., Gurca, A., Brunswicker, S.: Problem types and open innovation governance modes: a project-level empirical exploration. IEEE Trans. Eng. Manage. (2019). https://doi.org/10.1109/TEM.2019.2942132
4. Culpan, R.: Global business alliances. In: Theory and Practice. Quorum Books, Westport, Conn. (2002)
5. Sydow, J., Braun, T.: Projects as temporary organizations: an agenda for further theorizing the interorganizational dimension. Int. J. Project Manage. (2018). https://doi.org/10.1016/j.ijproman.2017.04.012
6. vom Brocke, J., Lippe, S.: Managing collaborative research projects: a synthesis of project management literature and directives for future research. Int. J. Project Manage. (2015). https://doi.org/10.1016/j.ijproman.2015.02.001

7. Kuster, J., Bachmann, C., Huber, E., Hubmann, M., Lippmann, R., Schneider, E., Schneider, P., Witschi, U., Wüst, R.: Handbuch Projektmanagement. Agil—klassisch—hybrid, 4th edn. Springer Gabler, Berlin (2019)
8. Dieterich, K., Kaschub, V.L., Ohlhausen, P.: Ein projektkulturbewusstes Management-Modell für interorganisationale F&E-Projekte. In: Binz, H., Bertsche, B., Spath, D., Roth, D. (eds.) Stuttgarter Symposium für Produktentwicklung SSP 2021. Konferenz zum 6. Stuttgarter Symposium 2021, pp. 1–12 (2021)
9. Mentimeter. https://www.mentimeter.com/ (2021)
10. Haritz, A.: Innovationsnetzwerke. Ein systemorientierter Ansatz. Zugl.: Mainz, Univ., Diss., 2000 u.d.T.: Haritz, André: Heterarchische Innovationsnetzwerke: eine heuristische Konstruktion auf Basis des Systemansatzes, 1st edn. Gabler Edition Wissenschaft Forum produktionswirtschaftliche Forschung. Dt. Univ.-Verl., Wiesbaden (2000)
11. Sydow, J.: Management von Netzwerkorganisationen—Zum Stand der Forschung. In: Sydow, J. (ed.) Management von Netzwerkorganisationen, pp. 373–470. Gabler Verlag, Wiesbaden (2010)
12. Frey, D., Henninger, M., Lübke, R., Kluge, A.: Einführung und konzeptionelle Klärung. In: Frey, D. (ed.) Psychologie der Werte. Von Achtsamkeit bis Zivilcourage—Basiswissen aus Psychologie und Philosophie, pp. 1–12. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
13. Butler, J.K.: Toward understanding and measuring conditions of trust: evolution of a conditions of trust inventory. J. Manag. (1991). https://doi.org/10.1177/014920639101700307
14. Schön, W.: Vertrauensorientiertes Projektmanagement. Top-10-Erfolgsfaktoren für Projekte und Veränderungsprozesse, 1st edn. essentials. Springer Fachmedien Wiesbaden; Imprint: Springer Gabler, Wiesbaden (2020)
15. Rieger, V.: Kultur und innovationen. Empirische Studien auf Ebene von Ländern, Organisationen und Teams. Dissertation, Research. Springer Fachmedien Wiesbaden GmbH (2018)

Using an Actual Design Method for the Design of Research Methodologies: Case of the Dichotomy Exploration and Exploitation in Context of Innovation Management

Rob Dekkers and Qijun Zhou

Abstract Whereas scholarly studies often elaborate on the design of the method for empirical data collection and analysis, surprisingly, the actual use of design principles taken from the domain of design and engineering of products has not come into the picture, yet; the stance of this paper is that the use of design principles taken from this domain may be beneficial for making choices and detailing research design. The application is demonstrated through showcasing the research design process of a doctoral study. The topic explored in this doctoral study concerns the theoretical concepts of exploration and exploitation, especially focused on their implication in managerial practices of managing innovation. In addition, both theoretical and empirical validity of view exploration and exploitation as a dichotomy was questioned; this was done through challenging the domain assumptions sustaining these notions. In order to tackle this rather complex research problem, the study adopted Pugh's 'controlled convergence method' to have a more systematic approach in setting the final research methods.

Keywords Controlled convergence method · Design and engineering · Research methodology

1 Introduction

Whereas scholarly studies often elaborate on the design of the method for empirical data collection and analysis, surprisingly, the actual use of design principles taken from the domain of design and engineering for artefacts has not come into the picture, yet. When guidance is provided about the design of the research methodology it is often framed as selection of a 'study design', aka research design and research

R. Dekkers (✉)
University of Glasgow, Glasgow, Scotland
e-mail: rob.dekkers@glasgow.ac.uk

Q. Zhou
University of Greenwich, London, UK

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_71

methodology. In the discipline of design and engineering this terminology is not used in this manner. There is referral to developing solutions or concepts for artefacts, which means going stage-wise from eliciting requirements to detailed instructions for use. It carries some resemblance to study design in terms of starting with broadly formulated research questions or hypotheses as an outcome of a literature review to developing the protocol for actual data collection and data analysis. This analogy made us wonder whether it would be beneficial to apply principles from design and engineering to the design of research methodology.

1.1 Background and Objectives

Particularly, for novices to research, such as doctoral, postgraduate and early career researchers, the design of a research methodology may be daunting, while a justification is required, too. As Groenewald [1: 42] states the selection of the most appropriate design of an empirical study requires knowledge of a vast range of methods; this reverberates in the words of Clarke and Visser [2: 456], too, when they state: 'Without a full understanding of the range of methodologies, and their suitability to answer specific research questions—all of which may seem apparent to an experienced researcher—novice, or early career researcher-explorers may sometimes feel lost …' Consequently, Groenewald [1: 42] states that selecting an appropriate research design is cumbersome for these scholars. In the same vein, the commentary by Vermunt [3: 332–3] refers to challenges exacerbated by methodological camps and differences across educational systems. Setting these debates aside, since a study design normally is seen as the choice for a specific research method, the literature abounds with advice for specific research methods. Only when methods are combined, with the combination of qualitative and quantitative methods being a prominent case in point, there is more attention to comparing methods. An example is the work by Johnson and Onwuegbuzie [4] when they discuss mixed-methods research, particularly focusing on putting together qualitative and quantitative methods in different configurations. Instigated by positive experiences from their doctoral degree, Jogulu and Pansiri [5] advocate mixed-methods studies, but they [5: 693–4, 695–6] hardly elaborate on the exact reasons for the design of the research. In such sources there is no reference to principles for design derived from design and engineering processes that may have informed the actual design of the research methodology.

The reason that principles and methods used for design and engineering are of interest has a number of reasons. When studies select or propagate a method, they often assume that the choice is a decision, often early in the conduct of the research project, or sometimes in advance due to reasons such as seeking research funding. Moreover, once a design is selected, all a researcher or research team has to do is finding more detail about the research method to actually undertake the empirical study. This differs from design and engineering processes that are teleological-oriented and iterative by nature. Also, a key feature of design and engineering is the elicitation of requirements, think about specifications for the design of a product or building, as a stepwise process of refinement. Differently, the selection of a study design is often limited to consideration of criteria, for example, Hulley et al. [6: 19] use FINER (feasible, interesting, novel, ethical, relevant) to that effect. Thus, taking a perspective of design and engineering, i.e. focusing on the elicitation of requirements, consideration of principal research methods and stepwise development of a research methodology as an integrated approach based on progressive insight, could yield additional benefits for design of research methodologies.

1.2 Scope and Outline of Paper

Therefore, the stance of this paper is that the use of design principles taken from the domain of design and engineering may be beneficial for making choices with respect to research design of empirical studies and detailing the research method. To this purpose, we will elaborate on principles of design and engineering. In the third section, one of the methods of design and engineering will be used to demonstrate the application to a research project. A final section concludes by dwelling on the benefits of the design and engineering approach.

2 Relevant Principles of Design and Engineering

As mentioned before, the teleological orientation of design and engineering processes for artefacts leads to three characteristics: elicitation of requirements, stage-wise development from concepts to detailed instructions and iterations to optimise the concept for the artefact in terms of performance. These three characteristics are discussed in the next section together with two main approaches and reflection on the analogy with design of research methodologies.

2.1 Three Characteristics for Design and Engineering of Artefacts

In terms of design and engineering approaches for artefacts, eliciting requirements is an essential step. Often, a distinction is made between so-called functional requirements, those that specify what an artefact establishes, and constraints that limit the use of materials and resources for either developing or producing the artefact. There are various ways that these lists of specifications are obtained. This can include end user involvement (for instance [7]), or investigating what are requirements. The latter is also found in requirements engineering; for example, Zoghwi and Coulin [8] provide an overview of these methods and tools in use for software engineering. Normally, these requirements and constraints may lead to trade-offs during the design and engineering process. Thus, the requirements and constraints form the starting point for the design and engineering of an artefact, though they may need to be revisited when detailing concepts.

A further characteristic of design and engineering approaches is that their processes are stage-wise organised and iterative in nature to achieve that an artefact meet its requirements. For example, Radhakrishnan and McAdams [9: 378] describe design and engineering activities as an iterative process consisting of phases, such as the generation of ideas, development of concepts, detailing of design and construction of prototypes. Also, Chang et al. [10: 275–7] allude to the iterative nature of design and engineering. A cause of these cycles is that information in earlier stages is often incomplete and inaccurate [11: 319], leading to revisiting earlier choices for the design of artefacts.

Typical for design and engineering is adapting principal solutions to turn them into feasible artefacts. This requires evaluating when detailing concepts to meet requirements. This process can lead to making trade-offs between conflicting requirements during the processes of design and engineering, which may occur during the successive stages. This leads to so-called secondary design and engineering processes [12: 321], which are also denoted as engineering change management; for an indicative overview of the latter, see Hamraz et al. [13]. Documented decisions for the design and trade-off will facilitate tracing the impact of changes that need to be made. Thus, design and engineering of artefacts is an iterative process to search for concepts that are feasible and incorporate changes for progressive insight.

Relevant to the quest in this paper, there are two main approaches with regard to principal solutions: point-based design and the controlled convergence method. Point-based design emphasises more detail on principal solutions in early stages through feasibility studies. The evaluation is based on as detailed requirements as possible. After selection of the most appropriate principal solution or conceptual design, a linear process of development is followed; however, commonly iterations and trade-offs will occur due to further detail becoming known, which cause reconsideration of trade-offs made in terms of feasibility, and this could lead to finding solutions within the chosen concept to accommodate functional criteria and constraints. Differently, the controlled convergence method, based on Pugh's [14] concept selection method, consists of the parallel processes of detailing requirements and developing concepts in stages; it has become more known as set-based concurrent engineering derived from the practices for new product development in the Toyota Production System [15: 913–4, 16: 48–49]. Subsequently in stages, more information about functional criteria, constraints and trade-offs become available. In parallel, the design and engineering activities move from conceptual design to detailed design. Because choices for the design are delayed, there is more opportunity for seeking optimal solutions for the artefact; the latter was noted earlier in this section as information becoming progressively available during the stages of design and engineering. As Raudberget and Sunnersjö [17: 368] note the approach may take more effort and lead-time, but this is set off by increased product success seen as innovative concepts, product performance and cost. Thus, there are two principal approaches to organising the process for design and engineering of artefacts—point-based design and controlled convergence method—that lead to differences in which stage the final concept is chosen, how (progressive) insight is used for making decisions with regard to advancing and discarding concepts, and how detailing of concepts takes place.

Taking the analogy between the design of artefacts and design of research methodology means equating concepts and principal solutions with research methods. The functional requirements are what needs to be known. Differently from the development of artefacts, what needs to be known is derived from literature, sometimes complemented by what stakeholders put forward (for example, in the case of pragmatic research). In addition to what needs to be known, constraints have to be met. Besides the temporal facet and resources for conducting the research project, practical considerations can extend to factors such as access to software for qualitative analysis. Looking at the two main approaches to design and engineering of artefacts, point-based engineering is closer to the canonical approach of selecting a method for empirical studies. However, applying this design and engineering approach to design of studies requires that methods are transferred into conceptual designs with some points detailed to assess feasibility; such requires the vast knowledge or the acquisition of this knowledge, as mentioned in the introductory section of this paper. In addition, for this approach to work the full requirements for the selection of an appropriate research method need to be elicited and detailed. A setback of this approach is that later identified bottlenecks or trade-offs are set within a research method and this may restrict data collection and analysis, which are normally found as limitations of a study. An example would be selecting the case study methodology and then finding out that variance across cases is too large to draw meaningful cross-case inferences. The controlled convergence method, the second main approach to design and engineering of artefacts, allows developing a more coherent design in stages. The advantage would be not being bogged down into much detail in earlier stages and seeking for more detailed information later when there is more progressive insight how the research can be conducted. However, this may come at the expense of more effort and longer lead time for the design of the research methodology. Thus, both main approaches for the design and engineering of artefacts can also be applied to the design of research methodologies, though point-based design in principle reflects the canonical approach; both methods—point-based design and controlled convergence method—have advantages and drawbacks.

2.2 Process Model for Design of Research Methodology

Applying the controlled convergence method to the design of the research methodology with the purpose of addressing the research questions has led to the creation of a process model. Generally, this process model includes three key phases; see Fig. 1. The first stage sets the 'foundation' of the empirical study. Based on the aim and objectives of a study, this stage will identify (i) the level of analysis in the study, (ii)

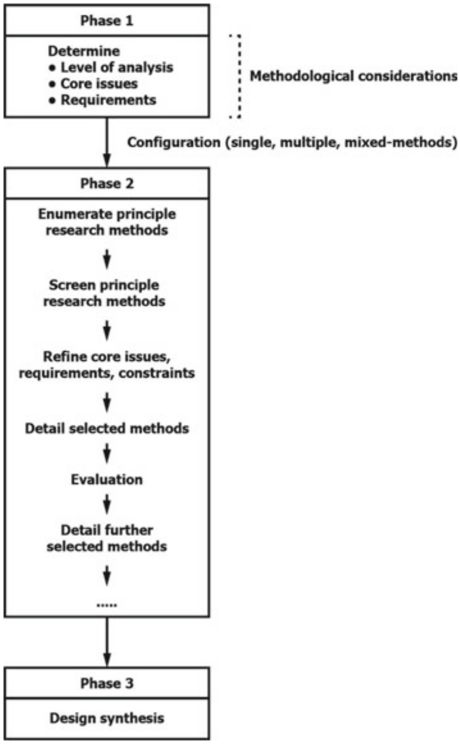


Fig. 1 Generic process for design of research methodology

core issues that need to be addressed and (iii) functional requirements and constraints. This process is akin to eliciting requirements in the case of design and engineering for artefacts. Consideration of these three points of departure will lead to the decision whether the research will be a single, multiple methods or mixed-methods study. The second stage focuses on the actual method selection and evaluation, following the controlled convergence method. This means that the refinement of core issues, requirements and constraints may happen in few steps. The third stage is design synthesis, in which the elements of the design of the research methodology are amalgamated. This leads to the protocol for the empirical study as the final outcome. The process, as displayed in Fig. 1, follows the principles of the controlled convergence method, and leads to a stepwise development and design of the research methodology detailed in a protocol.

3 Application to Design of Research Methodology

The principles of the controlled convergence method were applied to the design of an empirical study into the dichotomy of exploration and exploitation, introduced in the seminal paper of March [18], with a focus on the domain of innovation management. A literature review led to the overall conclusion that this dichotomy, even though widely accepted, contains fundamental concerns with regard to its theoretical foundations and is not necessarily rooted in actual managerial practices. The question was how to design the research methodology to most effectively find out about these theoretical foundations and what can be observed. However, selecting a method in the traditional way or by using 'point-based design' would not yield insight from multiple perspectives. Writings, such as Johnson and Onwuegbuzie [4] about mixed-methods studies, could lead to some insight, but leave open actual choices and how to detail the research method. Given the lack of precedence in literature in terms of challenging the dichotomy, the complexity of the research design and the need to develop a coherent design for the empirical study, led to taking the controlled convergence method as starting point.

3.1 Brief Overview of Outcomes from Literature Review

A literature review mainly revealed three points of contention in the current literature regarding the study of exploration and exploitation in the context of innovation management. First, the conceptualisation of exploration and exploitation is used inconsistently across studies, leading to confusions in treating them as a dichotomy. Some publications put the emphasis when measuring exploration and exploitation on activities or processes, for example [19], whereas others view them as outcomes. for instance [20]. Second, viewing exploration and exploitation as a dichotomy has not been challenged in current studies, even though the related concept of ambidexterity is built on the premise that exploration and exploitation co-exist in organisations. Third, there is lack of direct empirical evidence supporting the relevance of this theoretical dichotomy for managerial practice. One main cause for this is that most studies have used the disjunction between exploration and exploitation for their theoretical lens. This is reflected in the research methods of studies into innovation management, building on the dichotomy. It appears that quantitative studies dominate, for which the disjunction results in present variables that are attributed to either exploration or exploitation. Such a study design will not lead to disproving the distinction. However, the qualitative studies have also taken the dichotomy as starting point, with an example being the case study by Cantarello et al. [21]. Therefore, appraising scholarly writings resulted in highlighting inconsistencies how the dichotomy is measured, implicit evidence that exploration and exploitation co-exist by default (thus, not being extremes as suggested by the contrasting notion), bias towards the dichotomy by its operationalisation in studies and lacking evidence of its utilisation in practice.

These outcomes of the literature review have motivated the study to make two arguments regarding exploration and exploitation. The first argument here is that when scrutinising the definition of exploration and exploitation in studies it appears that these concepts can only be identified based on outcomes of activities or processes.

Thus, if exploration and exploitation can only be defined based on outcomes of activities or processes, then the dichotomy is only helpful in evaluating past performance, while not being used for managing innovation because it has not necessarily impact on managerial decision making. The second argument is that the lack of both theoretical challenges to the dichotomy as well as no direct empirical evidence, it may be the case that there is no 'pure' exploration or 'pure' exploitation in managerial practice. This is to say that exploration and exploitation cannot be defined or occur without each other, for which the rise of introducing ambidexterity for the dichotomy is an indirect conjecture. Consequently, it will be up to an empirical investigation to shed lights on these two arguments.

Furthermore, this empirical investigation is led by four research questions derived from the literature review and the two arguments:

- **Research Question 1**: How to identify organisations that manage exploration and exploitation efficiently?
- **Research Question 2**: How does performance in terms of outcomes and resource allocation in the context of ambidexterity influence recurrent decision-making for innovation management?
- **Research Question 3**: What are practices for exploration and exploitation in managing innovation, and are they treated separately in practice to the extent assumed by March [18]?
- **Research Question 4**: How could innovation management be understood if exploration and exploitation are not separated in practice?

3.2 First Phase for Design of Research Methodology

Based on the literature review, evaluation of previous research designs adopted by studies and research questions for the doctoral study, first core issues, functional requirements, constraints and level of analysis are identified. Here, core issues can be understood as the ultimate target this study wishes to accomplish through its design of the research methodology, whereas functional requirements refer to what needs to be established and constraints to practicality. With regard to the level of analysis, especially, the nature of the first research question requires analysis of the relative performance of firms in industrial settings. This can only be achieved through quantitative analysis at industrial level, i.e. analysis of performance in industrial sectors. In this respect, there are only few studies that looked at industrial level, for example [22]. Addressing the other three research questions will rely more on investigating internal processes and decision making with regard to innovation management; since the dichotomy is challenged this should be undertaken without the preconceived separation of exploration and exploitation, as embodied by the dichotomy. Thus, it is more appropriate to follow a qualitative approach for data collection and analysis at the firm level. Moreover, studies from the literature review that have adopted an analysis at both industrial and firm level are limited to examining the impact of external

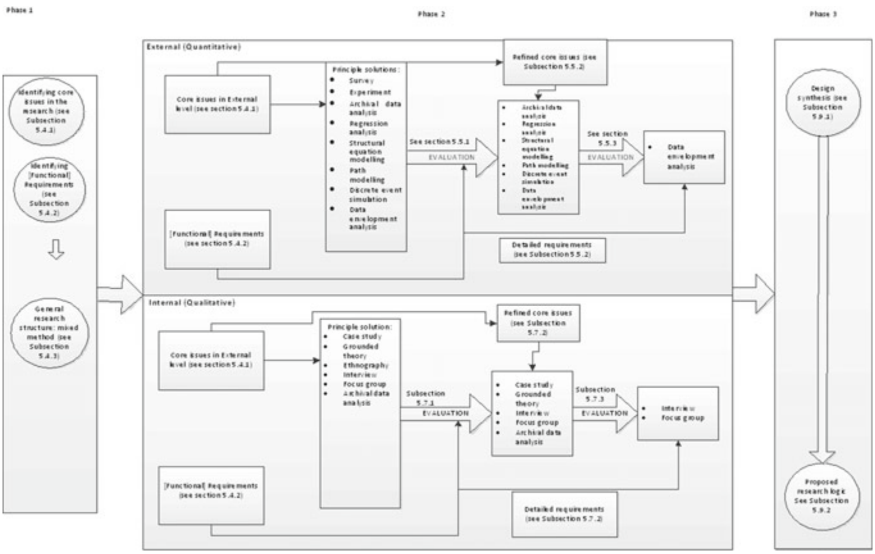


Fig. 2 Actual design of research methodology

factors on managing exploration and exploitation, especially by adopting survey-based designs; for instance [23]. Combining both levels without prejudice towards the dichotomy would result in a study making a contribution to scholarly knowledge. This means that for the doctoral study, it is appropriate to adopt a mixed-methods approach, with a quantitative analysis at the industrial level complemented by a qualitative approach at firm level, thus ensuring a unique contribution to theoretical foundations for innovation management.

3.3 Design of Research Methodology

The resulting design of the methodology for this doctoral study is found in Fig. 2. This figure displays the process and the decisions made during the design of the study.

4 Concluding Remarks

As highlighted by the case, there are advantages of using the design and engineering approach. First, the design and engineering approach leads to a more intricate design of the research methodology, particularly when there are multiple points for the empirical study following from the literature review. The controlled convergence method facilitates researchers to go in a systematic manner from a coarse analysis of research methods to a more granular review of research methods. Using progressive insight into advantages of specific methods and the development of criteria allows the focus to remain on what the research methodology should resolve in terms of original research objectives. Second, the balanced consideration of research methods and how to detail these for the actual data collection and analysis, results in a design of the research methodology that can be better justified. Were the canonical approach followed, most likely the case study methodology would have been opted for. This would have resulted in less directional collection of data and fewer opportunities for triangulation. This implies that the traditional approach—selecting a research method—possibly relies more on knowledge and experience with regard to specific research methods, practices for research methods in domains and preferences related to research paradigms. Third, the design and engineering approach connects the design of the research methodology better to the literature review. It takes as starting point what is required for the data collection and subsequent analysis. This calls on an adequate literature review to determine the state of scholarly knowledge and how a study builds on it. Fourth, practical concerns are taken into account as constraints. Fifth, it allows to systematically consider changes that happen during the conduct of an empirical study. Sixth, during the doctoral viva voce no concerns were raised about the design of the research methodology, though a common question for these examinations. These six advantages demonstrate it is worthwhile to consider this alternative approach to the design of research methodologies rather than following more canonical approaches of selecting a research method early on.

References

1. Groenewald, T.: A phenomenological research design illustrated. Int. J. Qual. Methods **3**(1), 42–55 (2004). https://doi.org/10.1177/160940690400300104
2. Clarke, E., Visser, J.: Pragmatic research methodology in education: possibilities and pitfalls. Int. J. Res. Method Educ. **42**(5), 455–469 (2019). https://doi.org/10.1080/1743727X.2018.1524866
3. Vermunt, J.D.: Conceptions of research and methodology learning: a commentary on the special issue. Scand. J. Educ. Res. **49**(3), 329–334 (2005). https://doi.org/10.1080/00313830500109659
4. Johnson, R.B., Onwuegbuzie, A.J.: Mixed methods research: a research paradigm whose time has come. Educ. Res. **33**(7), 14–26 (2004). https://doi.org/10.3102/0013189X033007014
5. Jogulu, U.D., Pansiri, J.: Mixed methods: a research design for management doctoral dissertations. Manag. Res. Rev. **34**(6), 687–701 (2011). https://doi.org/10.1108/01409171111136211
6. Hulley, S.B., Cummings, S.R., Browner, W.S., et al.: Designing Clinical Research. Lippincott Williams & Wilkins, Philadelphia (2013)
7. Kaulio, M.A.: Customer, consumer and user involvement in product development: a framework and a review of selected methods. Total Qual. Manag. **9**(1), 141–149 (1998). https://doi.org/10.1080/0954412989333
8. Zowghi, D., Coulin, C.: Requirements elicitation: a survey of techniques, approaches, and tools. In: Aurum, A., Wohlin, C. (eds.) Engineering and Managing Software Requirements, pp. 19–46. Springer, Berlin (2005)

9. Radhakrishan, R., McAdams, D.A.: A methodology for model selection in engineering design. J. Mech. Des. **127**(3), 378–387 (2005). https://doi.org/10.1115/1.1830048.0
10. Chang, A.S.T., Shih, J.S., Choo, Y.S.: Reasons and costs for design change during production. J. Eng. Des. **22**(4), 275–289 (2011). https://doi.org/10.1080/09544820903425218
11. Shishank, S., Dekkers, R.: Outsourcing: decision-making methods and criteria during design and engineering. Prod. Plan. Control Manag. Oper. **24**(4–5), 318–336 (2013). https://doi.org/10.1080/09537287.2011.648544
12. Dekkers, R., Chang, C.M., Kreutzfeldt, J.: The interface between "product design and engineering" and manufacturing: a review of the literature and empirical evidence. Int. J. Prod. Econ. **144**(1), 316–333 (2013). https://doi.org/10.1016/j.ijpe.2013.02.020
13. Hamraz, B., Caldwell, N.H.M., Clarkson, P.J.: A holistic categorization framework for literature on engineering change management. Syst. Eng. **16**(4), 473–505 (2013). https://doi.org/10.1002/sys.21244
14. Pugh, S.: Concept selection: a method that works. In: Hubka, V. (ed.) International Conference on Engineering Design, pp. 497–506. Rome (1981)
15. Salgado, E.G., Dekkers, R.: Lean product development: nothing new under the sun? Int. J. Manag. Rev. **20**(4), 903–933 (2018). https://doi.org/10.1111/ijmr.12169
16. Ward, A., Liker, J.K., Cristiano, J.J., et al.: The second Toyota paradox: how delaying decisions can make better cars faster. Sloan Manag. Rev. **36**(3), 43–61 (1995)
17. Raudberget, D., Sunnersjö, S.: Experiences of set based concurrent engineering in four product developing companies. In: Horváth, I., Mandorli, F., Rusák, Z. (eds.) 8th International Symposium on Tools and Methods of Competitive Engineering, pp. 357–370. Ancona (2010)
18. March, J.G.: Exploration and exploitation in organizational learning. Organ. Sci. **2**(1), 71–87 (1991). https://doi.org/10.1287/orsc.2.1.71
19. Atuahene-Gima, K.: Resolving the capability-rigidity paradox in new product innovation. J. Mark. **69**(4), 61–83 (2005). https://doi.org/10.1509/jmkg.2005.69.4.61
20. Bierly, P.E., III., Damanpour, F., Santoro, M.D.: The application of external knowledge: organizational conditions for exploration and exploitation. J. Manage. Stud. **46**(3), 481–509 (2009). https://doi.org/10.1111/j.1467-6486.2009.00829.x
21. Cantarello, S., Martini, A., Nosella, A.: A multi-level model for organizational ambidexterity in the search phase of the innovation process. Creativity Innov. Manage. **21**(1), 28–48 (2012). https://doi.org/10.1111/j.1467-8691.2012.00624.x
22. Gilsing, V., Nooteboom, B.: Exploration and exploitation in innovation systems: the case of pharmaceutical biotechnology. Res. Policy **35**(1), 1–23 (2006). https://doi.org/10.1016/j.respol.2005.06.007
23. Yang, T.T., Li, C.R.: Competence exploration and exploitation in new product development. Manag. Decis. **49**(9), 1444–1470 (2011). https://doi.org/10.1108/00251741111173934

Developing High-Variant Products and Production Systems in Line with Value Stream Requirements Through Simulation

Oliver Riedel, Dirk Marrenbach, and Oliver Scholtz

Abstract Markets are increasingly demanding greater product variety and, at the same time, ever shorter delivery times. In the case of new product development, it is often no longer possible for production to ensure a sufficiently high delivery capability with the help of lean methods and value stream methods. Especially if acceptable inventories and costs are assumed. During product development, products and product systems have not yet been sufficiently coordinated with each other from a value stream perspective and designed in a targeted manner. For this reason, it is crucial to align the product and its future production system with the goal of a functioning value stream as early as the concept phase of product development. Up to now, the value stream methodology has been used to analyze and optimize real production processes that are already running. In the future, the delivery times and delivery reliability of new products are to be secured by simulating the value stream based on the value stream methodology. In the context of Advanced Systems Engineering (ASE), the value stream resulting from the product is already created in an early phase of product development in order to be able to check the functionality of the production system and the deliverability of the product. Value stream simulation makes a significant contribution to improving cooperation between product development, production and logistics, as we can vividly present in this article using the ASE demonstration product. The value stream simulation revealed that the demonstration product had design weaknesses in terms of the required delivery capability. Assuming that no changes to the pro-duction system are possible, design changes were made to the product. The housing, which runs through a bottleneck line with over 100 variants, was standardized and the number of variants reduced to two. This led to later variant creation in the value stream. The variants now only appear in assembly, which can cope with many variants due to its high variant flexibility. The

O. Riedel · D. Marrenbach · O. Scholtz (✉)
Fraunhofer Institute for Industrial Engineering, Stuttgart, Germany
e-mail: oliver.scholtz@iao.fraunhofer.de

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_72

simulation of the revised product and the new value stream resulted in a significant improvement in delivery capability and the required short delivery times to the customer can be met.

Keywords Product design · Value stream simulation · Deliverability

1 Introduction

1.1 The Challenge of Variant Diversity and Short Delivery Times for Production Companies

Globalization, liberalization, digitalization demographic change and climate change force a dramatic transformation in market. Business, working and living conditions. growing volatility, uncertainty, complexity and ambiguity are the characteristic features of the surrounding in which production systems had to survive. Agility, flexibility and traceability are the key features for value stream systems [1, 2].

The principles, strategies and methods of lean management improve the development of flexible as well as efficient productive and rentable production and logistical systems [3]. Each industrial branch uses today these well known, well established and proven methods to optimize their systems and processes [4–6]. Each market competitor offers now products and services with nearly identical functionality, quality and costs using nearly the same process design. Today the offer of high-quality products and services in combination with minimal costs are expected by all costumers. Under these competitive market conditions effective processes and systems are an established standard in all industrial branches. In this situation the delivery time and the delivery quality become the most important unique selling point.

In this competitive market situation, the development of individual designed products and services are the last possibility to survive. The development and production of individual designed products rise the dynamic complexity in the whole value stream system. The number of products as well as the number of product variations grow because of the wide spectrum of individual needs.

1.2 Value Stream Design as a Solution Approach for the Coordination of Product and Production System

In traditional organized value streams, the needed capability to deliver and the delivery quality is guaranteed through the increasing of the inventory. The ability of deliver rise because every product variation is always available. The inventory costs rise dramatically because of the high level of unordered products.

In this situation the proven principles, strategies and methods of lean management are no longer able to balance the inventory level and the capability of deliver by a flow-oriented design of the whole value stream [3, 7]. The additional instruments of Value Stream Engineering are the toolset to reach this goal. In the center of the Value Stream Engineering stands the analysis and modelling approach value stream map [2]. In a value stream map, the order, material and information flows for the manufacture of a product or a product family are mapped with the help of a process-oriented modeling approach [8]. The basis for the value stream map is a chain of processes that a product goes through during its production from incoming goods to outgoing goods. The process chain is made up of a sequence of processing, checking, picking and packaging processes that are linked via material, information and order flows. Storage, buffers and storage facilities can also be integrated into the material flows. The processes are coordinated either directly between the individual process modules (Kanban control) or indirectly via the central order control (PPS system). Each process and each stream are assigned with a set of key performance indicators that allow the calculation of the delivery time. The "value stream map" modeling approach is based on a standardized collection of symbols, key figures and calculation rules.

The value stream map impresses with its simple, easy-to-understand structure as well as its ease of learning and use. In the last decades the value stream engineering approach has been widely used to optimize the production system as well as the whole supply chain.

The discussed problems with the delivery time result on a lack of integration between product development, production and logistics [9]. Short delivery times can be reached through a continuous optimization process between product development, production and logistics. The capability to deliver can be optimized by a flow friendly product design accepting the possibilities and limitations of the whole value stream system.

In this article we present an integral approach for the simulation-based development of products and processes based on the value stream map to minimize the gap between product development, production and logistics.

2 Integration of Value Stream Engineering and Advanced Systems Engineering

2.1 Definition of Advanced Systems Engineering

Systems Engineering deals with the holistic development, realization and optimization of complex systems in industry, military, economy, science, state and society. For this purpose, systems engineering provides a collection of principles, strategies, models, documents, methods and procedures. The ISO 15288 defines a system as "A combination of interacting elements organized to achieve one or more stated purposes". Advanced systems Engineering (ASE) is a further development of the systems engineering using the possibilities of a fully digitalized system life cycle. Advanced Systems Engineering is a collection of principles, strategies and methods to develop and realize complex products and processes using a model driven development methodology. The aim of Advanced Systems Engineering is the systematic usage of digital tools to improve cross-disciplinary collaboration not only in companies.

2.2 Definition of Value Stream Mapping (VSM)

The value stream mapping is a static model that cannot take into account the influence of dynamic and stochastic effects on the capability to deliver. In its original form, the VSM is a simple graphic model that is created and parameterized manually during a factory tour [2]. A value stream map presents only the process chain of a single product or a single product family. Normally numerous products or product families have to share resources flowing through the value stream network. Sharing resources increase the unplanned grow of the delivery time caused by scarce capacities as well as the possibility of delays caused by product changes. If the number of variants is high, the whole supply chain will play an increasing role. Accordingly, the suppliers must be fully integrated into the analysis and their delivery times must be integrated. For this purpose, the traditional Value Stream Map has been expanded to include suppliers and transports. In this context, the Extended Value Stream Map forms the basis for developing a conceptual model of the entire value-added network or a product family as well as for simulating the value stream network.

2.3 Definition of Simulation

Simulation is the reproduction of a system with its dynamic processes in an experimental model in order to obtain knowledge that can be transferred to reality. With the help of experimental models, simulation technology analyzes systems. The results of the simulation experiments are used to design and optimize the systems under consideration. This procedure for planning and optimizing systems is described in VDI 3633 [10] as a control loop of simulation technology. The systematic application of the simulation technology on a planning task includes the following steps [11]:

1. Define the tasks, requirements and goals of the simulation study.
2. Analyze the simulation field.
3. Develop, and validate the conceptional model.
4. Develop and validate the simulation model.
5. Plan, run and evaluate the simulation experiments.
6. Document the simulation experiments.
7. Interpret the simulation results.

2.4 Value Stream Simulation of Product and Production System to Ensure High Delivery Capability

Increasing complexity, dynamics and turbulences forces hierarchical and functional organizations in industry to overcome their traditional functional worldview and develop a new holistic integral worldview on their products, services and processes. The integration of Value Stream Engineering into the ASE is a step to develop an integral platform for the development products, services and processes using a digital modelling and simulation system.

Simulation-based Value Stream Engineering combines the steps for designing production systems based on the Static Extended Value Stream Map with the steps of a dynamic simulation-based design of Value Stream Networks [4, 6, 12].

The capability to deliver as well as the delivery time is the main result of the simulation study. The result of the simulation study gives detailed information under which conditions a new product can be produced. The capability of delivery has been proven under static and dynamic conditions. The simulation study offers the borders of the existing production and logistical system in relation to the designed product structures. The management can use the product-process-matrix, the process-resources-matrix, the value stream map and the value stream simulation model to improve the modularization of their products, services and processes. The systematic identification of vulnerabilities can be improved using these static and dynamic models.

The combination of static and dynamic design of value stream systems includes the following steps:

1. Define the product structure for a single product or product family (product structure matrix).
2. Define the process structure for the single product or product family (product process matrix).
3. Integrate the resource data into the process structure.
4. Develop the actual value stream map including external suppliers and transports.
5. Use the design principles to improve the actual value stream system.
6. Develop the future value stream map including the well-known production and logistics restrictions.
7. Integrate the future value stream map into the conceptional model for the whole value stream net.
8. Develop a simulation model for the value stream network.
9. Improve the single product value stream system as well as the value stream network.

2.5 Value Stream Compatible Design of Product and Production System

It is important to design a functioning value stream already in the development phase of a product, with short and reliable delivery times to the customer despite a high number of variants and large fluctuations in customer orders. The product and the production system must be coordinated in such a way that a functioning value stream is created [13].

Two essential characteristics have to be coordinated. In the case of the product, it is important to consider the number of variants of expensive parts, assemblies and products, which continue to increase in the course of the value stream. The increasing number of variants places a demand on the production system. The variants must always be reliably provided by the production system. The production system focuses on the property of variant flexibility. It enables the production system to produce the required variants of parts, assemblies and products quickly and flexibly. Variant flexibility is higher the smaller the setup times, and thus the possible lot sizes [14] and the higher the capacity reserves. These two factors determine a significant part of variant flexibility. Ideally, the variant flexibility of the production system also increases in parallel with the increase in the number of variants in the value stream.

Product design determines the increase in the number of variants in the value stream. The increase can thus be designed quickly or slowly by the development department. The production system design can partly carry out the variant flexibility of the systems itself (e.g. setup flexible production systems), partly only together with the product development (e.g. design of production processes with short setup times) and thus design sufficiently flexible for a high number of variants.

Due to these dependencies, the product and its production system must be jointly designed and coordinated according to the following basic principles:

1. **Value stream compatible product design: late differentiation**

In late differentiation, products are designed in such a way that the number of variants of parts, assemblies and products increases late in the value stream [15, 16]. This process strategy in dealing with variants is also referred to as postponement [15].

2. **Value stream-oriented production system design: design of variant-flexible processes** [16]

These are production systems designed according to lean principles with short setup times, small order batches and with sufficient capacity reserves to be able to absorb order peaks from customer orders or simultaneous requirements for several parts and assemblies [14].

The basic idea here is to allow variants to emerge only at such a late stage that the production system has sufficient variant flexibility in the value-adding stages. If the product and production system are coordinated, the number of variants increases only slowly in the value stream. Expensive parts that can only be manufactured with variant-intolerant production processes (e.g. casting) or on variant-intolerant production systems occur only in a few variants. Their variants only increase on production systems that have short setup times. Ideally, the variants only increase sharply in the last value-adding step, which has a high degree of variant flexibility with small setup times and a sufficient capacity reserve (e.g. manual assembly).

The functionality of the planned value stream and, if necessary, its improvements can be checked with a value stream simulation.

2.6 Example of a Value Stream Simulation and Value Stream Compliant Product Design

Figure 1 shows the principal procedure and the use of value stream simulation to ensure value stream-compliant product design, using the example of the ASE demonstrator product at Fraunhofer IAO. Advanced Systems Engineering methods are developed, tested and demonstrated on the ASE demonstrator (The methods are being developed as part of the "Cognitive Engineering and Production" project, which is funded by the Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.). Due to the high number of variants and its production on a bottleneck line, the lower part of the product represents the value stream-relevant part.

On the left side of Fig. 1, the initial situation, the bottom part is designed in 128 variants [64 drilling variants each with two blank variants (black and white blank)]. The variants differ in the drill and win holes, which are produced from the blank depending on the customer's choice of equipment. Due to the production of the lower part on a bottleneck line with long set-up times, the value stream simulation with the Plant Simulation© software tool results in a delivery capability of only 65%.

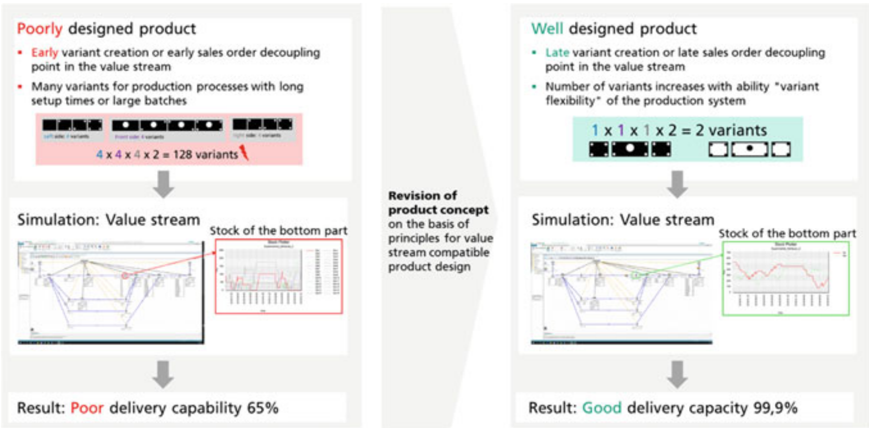


Fig. 1 Procedure for value stream-compliant product design by means of value stream simulation using the example of the ASE demonstrator product at Fraunhofer IAO

Due to market fluctuations, not all of the many bottom part variants can be produced in time despite the fact that they are kept in stock. Based on a product design of a gearbox manufacturer, the number of sub-part variants is reduced to two (black and white), as shown on the right side of Fig. 1. The two lower part variants receive all the drill holes that are required depending on the customer's equipment requirements. Unused drill holes are closed with plugs during final assembly. Variant formation thus takes place in the final value stream step, assembly. The bottleneck system with long setup times in production only has to produce two variants. The two variants can therefore always be produced at short notice if required. The value stream simulation results in a delivery capability of almost 100%.

3 Summary

The market demands more and more product variants and, at the same time, shorter delivery times. These conflicting requirements are increasingly difficult to represent in a functioning value stream. The risk that the existing production system cannot produce a newly developed product with its high number of variants in the delivery capability demanded by the market increases as a result.

Already in a very early phase of product development, the concept phase, the future functionality of the value stream can be checked or secured with a value stream simulation. Only a few data are required for the simulation, which are either already available or can be estimated by the developer or factory planner. The simulation is used to check whether the value stream is functional or whether the value stream or the factory can achieve the delivery capability required by the market.

If this is not achieved, then the value stream must be redesigned in such a way that the product and production system are better coordinated. To do this, either the production system must be enabled with regard to greater variant flexibility, or the product must be redesigned with regard to a more value stream-compliant design. Both designs are associated with costs. Future cost savings in ongoing operations are usually difficult to demonstrate in advance. Therefore, it is necessary to prove the necessity of the design measures by means of value stream simulations and to ensure that the planned measures are functional in the newly planned value stream.

In value stream design, it is important to find a holistic cost optimum between the two design objects of product and production system. In value stream design, improvements in both design objects must be considered, evaluated in terms of costs, checked for feasibility and a decision made as to which measures are to be implemented in the product and which in the production system.

References

1. Charron, R., James Harrington, H., Voehl, F., Wiggin, H.: The Lean Management Systems Handbook. CRC Press, Boca Raton (2014)
2. Erlach, K.: Value Stream Design—The Way Towards a Lean Factory. Springer, Berlin (2013)
3. Modrak, V.: Handbook of Research on Design and Management of Lean Production Systems. IGI Global (2014)
4. Jasti, N. V. K., Sharma, A.: Lean manufacturing implementation using value stream mapping as a tool: a case study from auto components industry. Int. J. Lean Six Sigma (2014)
5. McDonald, T., Van Aken, E., Rentes, A.: Utilising simulation to enhance value stream mapping: a manufacturing case application. Int. J. Logist. 213–232 (2002)
6. Güner Gören, H.: Value stream mapping and simulation for lean manufacturing: a case study in furniture industry. Pamukkale Univ. J. Eng. Sci. **23**, 462–469 (2017)
7. Klinker, R.: Schnellere prozesse allein helfen nicht. VDI nachrichten, Bd. Nr. 10 (2017)
8. Rother, M., Shook, J.: Sehen Lernen. LOG_X Verlag, Stuttgart (2000)
9. Abele, E., Brungs, F.: Durchlaufzeitverkürzung durch Produktoptimierung. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb, Nr. 105 (2010)
10. VDI3633: Simulation in Production, Material Handling and Logistics. Beuth Verlag, Düsseldorf (1993)
11. Marrenbach, D.: Simulationsgestützte Logistiksystemplanung auf Basis des Modells eines modular strukturierten, vollständigen technischen Anlagen-Entwurfs. Universität Stuttgart, Institut für Fördertechnik, Stuttgart (2008)
12. Lian, Y., Van Landeghem, H.: An application of simulation and value stream mapping in lean manufacturing. In: Proceedings 14th European Simulation Symposium, pp. 1–8 (2002)
13. Bley, H., Wuttke, C.: Logistikgerechte Produktentwicklung durch Kopplung von Produkt- und Produktionsprozeßmodellierung. In: Neue Generation von CAD/CAM-Systemen (VDI Berichte), Deutscher Ingenieur-Verlag (1997)
14. Balsliemke, F.: Mit kostenorientierter Wertstromplanung zu schlanker Logistik. In: Automobillogistik—Stand und Zukunftstrends. Springer Gabler, Wiesbaden (2016)
15. Gebhart, N., Kruse, M., Krause, D.: Gleichteile-, Modul- und Plattformstrategie. In: Handbuch Produktentwicklung, München, Carl Hanser Verlag (2016)
16. Göpfert, I., Schulz, M.: Strategien des Variantenmanagements als Bestandteil einer logistikgerechten Produktentwicklung - Eine Untersuchung am Beispiel der Automobilindustrie. In: Automobillogistik - Stand und Zukunftstrends. Springer Gabler, Wiesbaden (2016)

Quality Engineering and Management

Variable Sampling Plan Indexed by Taguchi's Quality Loss Under Emphasized Difference of Mean

Shuto Tanabashi, Ikuo Arizono, and Yasuhiko Takemoto

Abstract As an alternative to nonconforming product rate used traditionally as an index of quality, Taguchi has proposed quality loss. From the viewpoint of this new index of quality, some variable sampling inspection plans for assuring product quality indexed by quality loss have been proposed so far. The quality loss is defined as the expected loss by derivation from a target value of product quality characteristic. In concrete, the quality loss is composed of both of the square of the difference between the mean in the actual product quality and the target product quality, and the magnitude of variance in the actual product quality. And then, these two components have the same magnitude of effects in the evaluation of quality loss. On the other hand, it is known that for improving product quality at a production site, adjusting the mean in product quality is often easier than lessening the variance in product quality. Hence, we can see that adjusting the mean is more efficient and effective than lessening the variance as a way of reducing the quality loss. From the mentioned above, this research will consider a single sampling plan indexed by quality loss with consideration of emphasizing the loss derived from the mean in the product quality. And, the operating characteristics of the proposed sampling plan are confirmed through some numerical simulation.

Keywords Variable sampling plans · Taguchi's quality loss · Patnaik's approximation

1 Introduction

Sampling inspection is a useful method for assuring lot quality based on product quality characteristics from a part of a lot, when inspection is high cost or destructive.

S. Tanabashi (✉) · I. Arizono
Okayama University, Okayama, Japan
e-mail: tanabashi@s.okayama-u.ac.jp

Y. Takemoto
Kindai University, Osaka, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_73

There are different types of inspection depending on sampling procedures, and the most common inspection is single sampling. Nonconforming product rate, i.e., the proportion of nonconforming items, has been traditionally used as a measure of quality. However, it should be noted that since the proportion of nonconforming items means attribute property, information about variable property as observed values is lost. Therefore, even if individual observed values are different, lot quality may not be different under the measure of nonconforming product rate.

As an index of quality evaluated by variable quality characteristics, Taguchi [1] has proposed a quadratic form loss function based on the departure from a target value. Expected value of the quadratic form loss function has been named quality loss, and can be calculated using the mean and variance of the quality characteristic. The quality loss provides a more accurate and strict evaluation of the product quality. Based on the quality criterion as the quality loss, various sampling inspections have been considered by Arizono et al. [2], Yen and Chang [3], Aslam et al. [4] and Wu and Wang [5].

The quality loss is composed of both of the square of the difference between the mean in the product quality characteristic and the target product quality (called bias below), and the magnitude of variance in the product quality characteristic (called variation below). And then, these two components have the same magnitude of effects in the evaluation of quality loss. On the other hand, the bias may not be equal to variation from the viewpoint of the easiness in improvement on a production site. For example, reducing bias from the target value is sometimes easier than reducing variation of the product quality. In this case, improving the part of loss derived from the bias is one of the effective ways to improve the quality loss on the production site. As a sensitive method to detect the loss derived from bias, this paper examines a new single sampling inspection indexed by quality loss through emphasizing the loss by the difference between the mean in the product quality and the target value. Further, the performance of the proposed inspection is confirmed by simulation.

2 Single Sampling Plan Index by Quality Loss

Taguchi [1] has proposed the quadratic loss function which calculates product quality based on deviation from a target value. When the target value of the product quality is specified as T and the quality characteristic obeys the normal distribution $N(\mu, \sigma^2)$, the expected loss per item is formulated as follows:

$$k\tau^2 = E[k(x - T)^2] = k\{\sigma^2 + (\mu - T)^2\}, \tag{1}$$

where k denotes the proportional coefficient based on the functional limit of quality and the monetary loss brought by the product which cannot fulfill its function. Even if $k = 1$, statistical evaluation of quality does not lose generality. When $k = 1$, expected loss τ^2 is calculated from the sum of variance and square of deviation from the mean to the target value. In this paper, this is tentatively called quality loss.

Quality evaluation by quality loss may not be equal to evaluation by nonconforming production rate. Based on that, Arizono et al. [2] proposed a single sampling plan that uses quality loss as an index of quality. If the quality parameter (μ, σ^2) is unknown, the estimator of quality loss $\hat{\tau}^2$ is calculated from the random sample as follows:

$$\hat{\tau}^2 = s^2 + (\overline{x} - T)^2, \tag{2}$$

where s^2 is the sample variance and $\overline{x}$ is the sample average, and these are the maximum likelihood estimators of the variance and mean of the product quality. The inspection defines target quality as two types of quality losses: acceptance loss τ_0^2 and non-acceptance loss τ_1^2. The sample size and inspection criterion have been derived from an approximate model of the probability distribution of quality loss using the chi-squared distribution. The detail can be seen in Arizono et al. [2].

3 Single Sampling Plan Index by Quality Loss

As mentioned in the introduction, focusing on the loss $(\overline{x} - T)^2$ by the difference between the average and the target value is one of the effective ways to improve the quality loss on the production site. Based on this, a new sample statistic $\hat{\lambda}^2$ that emphasized the bias is defined as follows:

$$\hat{\lambda}^2 = s^2 + w(\overline{x} - T)^2, \tag{3}$$

where w is the coefficient that defines the ratio of the square of the bias to the variation. To emphasize the loss by the square of the bias, assume the relation of $w > 1$. If $w = 1$, the statistic $\hat{\lambda}^2$ is equal to the estimator of conventional quality loss $\hat{\tau}^2$. Note that $\overline{x}$ and s^2 are the maximum likelihood estimators of the mean and variance in the quality characteristic, which are calculated as follows:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{4}$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2, \tag{5}$$

where n is the sample size and x_i is a quality characteristic of each random sample from a lot.

Consider to employ the statistic $\hat{\lambda}^2$ as an inspection statistic to guarantee the quality loss τ^2. In concrete, a single sampling inspection with the operating characteristics under two criteria: acceptance quality and non-acceptance quality. In the proposed inspection, they are given as a quality loss: acceptance loss τ_0^2 and non-acceptance loss τ_1^2, where τ_0^2 is defined as the ideal quality of the product. Also, μ_0 and σ_0^2 are defined as a combination of the mean and variance that satisfies τ_0^2. The normal distribution $N(\mu_0, \sigma_0^2)$ is named the null distribution below. Since the ideal mean is equivalent to the target value, the acceptance mean μ_0 is equal to T. On one hand, the quality has variability. Therefore, the minimum possible variance on the production site is used as the acceptance and unavoidable variance σ_0^2. Therefore, the acceptance loss is calculated as $\tau_0^2 = \sigma_0^2$.

In the case of non-acceptance criteria, τ_1^2 shall be given from the product quality to be rejected. Then, the combination of (μ_1, σ_1^2) is defined as any combination of the mean and variance that satisfies τ_1^2. The normal distribution $N(\mu_1, \sigma_1^2)$ is named the alternative distribution below. There are innumerable pairs of (μ_1, σ_1^2).

We derived a single sampling inspection with the operating characteristics under acceptance quality and non-acceptance quality by employing $\hat{\lambda}^2$ as an inspection statistic. Sampling inspection by variables requires two values: sample size n and acceptance value of inspection statistic c. These values are required to satisfy the two kind of probabilities: producer's risk α and consumer's risk β. In order to calculate the probability of acceptance or non-acceptance, it is necessary to understand the statistical characteristics of the inspection statistic $\hat{\lambda}^2$.

First, consider the statistical model of $\hat{\lambda}^2$ under the acceptance loss τ_0^2. If the product quality follows $N(\mu_0, \sigma_0^2)$, statistic η_0 based on inspection statistic $\hat{\lambda}^2$ is defined as follows:

$$\eta_0 = \frac{n\hat{\lambda}^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2} + w\frac{n(\overline{x} - \mu_0)^2}{\sigma_0^2} \sim \chi_{n-1}^2 + w\chi_1^2, \tag{6}$$

where χ_ν^2 denotes chi-squared distribution with ν degrees of freedom. Since η_0 is given as a weighted sum of two chi-squared distributions, it is too complex to analyze for the statistical characteristic of statistic η_0. To represent the statistical characteristic of η_0 as a simple distribution, Patnaik's approximation [6] is applied to Eq. (6). In this method, the following new statistic ρ_0 is defined using η_0.

$$\rho_0 = \frac{2E[\eta_0]}{V[\eta_0]}\eta_0 \sim \frac{n-1+w}{n-1+w^2}(\chi_{n-1}^2 + w\chi_1^2), \tag{7}$$

where $E[\eta_i]$ and $V[\eta_i]$ means the expectation and variance of η_i respectively. The expectation and variance of ρ_0 are equal to those of the chi-squared distribution with ϕ degrees of freedom.

$$\phi = \frac{(n-1+w)^2}{n-1+w^2}. \tag{8}$$

Therefore, under the null distribution $N(\mu_0, \sigma_0^2)$, $\hat{\lambda}^2$ can be approximated using a single chi-squared distribution that satisfies the first and second order moments as follows:

$$\hat{\lambda}^2 = \frac{\sigma_0^2}{n}\frac{V[\eta_0]}{2E[\eta_0]}\rho \sim \frac{\sigma_0^2(n-1+w)}{n\phi}\chi_\phi^2. \tag{9}$$

Then, consider the statistical model of $\hat{\lambda}^2$ when the product quality obeys the alternative distribution $N(\mu_1, \sigma_1^2)$. Then, the statistical model of the statistic η_1 under the alternative distribution $N(\mu_1, \sigma_1^2)$ is defined as follows:

$$\eta_1 = \frac{n\hat{\lambda}^2}{\sigma_0^2} = \frac{\sigma^2}{\sigma_0^2}\left(\frac{ns^2}{\sigma^2} + w\frac{n(\overline{x}-\mu_0)^2}{\sigma^2}\right) \sim \frac{\sigma^2}{\sigma_0^2}(\chi_{n-1}^2 + w\chi_{1,\delta}^2), \tag{10}$$

where $\chi_{1,\delta}^2$ is the non-central chi-squared distribution with 1 degree of freedom and non-centrality parameter δ:

$$\delta = \frac{n(\mu_1-\mu_0)^2}{\sigma_1^2}. \tag{11}$$

To represent the statistical characteristic of η_1 as a central chi-squared distribution, Patnaik's approximation is applied to Eq. (10). The statistic ρ_1 is defined as follows:

$$\rho_1 = \frac{2E[\eta_1]}{V[\eta_1]}\eta_1 \sim \frac{n-1+w(1+\delta)}{n-1+w^2(1+2\delta)}\left(\chi_{n-1}^2 + w\chi_{1,\delta}^2\right). \tag{12}$$

Based on Eq. (12), ρ_1 can be approximated using chi-squared distribution with ψ degrees of freedom.

$$\psi = \frac{\{n-1+w(1+\delta)\}^2}{n-1+w^2(1+2\delta)}. \tag{13}$$

Therefore, under the alternative distribution $N(\mu_1, \sigma_1^2)$, $\hat{\lambda}^2$ is approximated using a single central chi-squared distribution with ψ degrees of freedom as follows:

$$\hat{\lambda}^2 = \frac{\sigma_0^2}{n}\frac{V[\eta_1]}{2E[\eta_1]}\rho_1 \sim \frac{\sigma^2\{n-1+w(1+\delta)\}}{\sigma_0^2\psi}\chi_\psi^2. \tag{14}$$

Inspection criteria are defined for each of the acceptance quality and non-acceptance quality. The acceptance criterion is derived from the approximate model based on the null distribution $N(\mu_0, \sigma_0^2)$, i.e., Eq. (9). In the case of the product quality following the acceptance loss τ_0^2, the probability of non-acceptance is defined as producer's risk α. Therefore, the acceptance criterion c_0 can be derived as follows:

$$c_0 = \frac{\sigma_0^2(n-1+w)}{n\phi}\chi_\phi^2(\alpha), \tag{15}$$

where $\chi^2_\nu(p)$ is the upper $100p$ percentile of the chi-squared distribution with ν degrees of freedom.

On the other hand, the criterion under the alternative distribution $N(\mu_1, \sigma_1^2)$ can be derived based on Eq. (14). In the case of the product quality following τ_1^2, the probability of non-acceptance is defined as $(1-\beta)$. Therefore, the non-acceptance criterion c_1 can be derived as follows:

$$c_1 = \frac{\sigma_1^2\{n-1+w(1+\delta)\}}{n\psi}\chi^2_\psi(1-\beta). \tag{16}$$

The sample size can be decided using acceptance and non-acceptance criteria. In the inspection, acceptance value must be less than non-acceptance value to satisfy both of the specified producer's and consumer's risks. Therefore, the following formula can be used to find the sample size.

$$\frac{\sigma_0^2(n-1+w)}{n\phi}\chi^2_\phi(\alpha) \leq \frac{\sigma_1^2\{n-1+w(1+\delta)\}}{n\psi}\chi^2_\psi(1-\beta). \tag{17}$$

However, the right term of Eq. (17) varies with the pair of (μ_1, σ_1^2). Hence, Eq. (17) needs to be satisfied in the case of all pair of (μ_1, σ_1^2). The inspection must set the value of the right term in Eq. (17) appropriately. Then, c_1 is defined as a lower limit value of non-acceptance. Therefore, c should be used as a minimum value as follows:

$$\frac{\sigma_0^2(n-1+w)}{n\phi}\chi^2_\phi(\alpha) \leq \min_{\mu_1,\sigma_1^2}\frac{\sigma_1^2\{n-1+w(1+\delta)\}}{n\psi}\chi^2_\psi(1-\beta) \equiv c. \tag{18}$$

Sample size n is decided as a minimum value satisfying Eq. (18).

4 Numerical Examples

In numerical examples, assume $T = 0.0$, $\tau_0^2 = 1.0$, $\tau_1^2 = 2.5$. The null distribution has been described as $N(0.0, 1.0)$, and the alternative distribution $N(\mu_1, \sigma_1^2)$ that satisfies the minimum value of c_1 has been searched by the numerical computation. Producer's and consumer's risks have been set as $\alpha = 0.05$ and $\beta = 0.10$.

First, the sample size and inspection criterion are verified. Figure 1 shows the relationship between sample size and coefficient w. When $w = 1.0$, this inspection is correspondent to the inspection proposed by Arizono et al. [2]. Figure 1 shows that the sample size is not constant as the coefficient increases. The range of minimum sample size is limited to $w \leq 1.80$ and the sample size is increased when $w > 1.85$. Figure 2 shows the shift of inspection criterion that calculated by Eq. (15). The inspection criterion increases with the coefficient w.

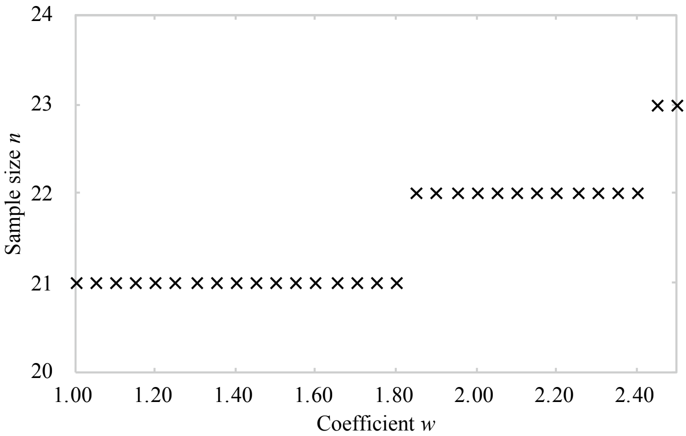


Fig. 1 Relationship between sample size n and coefficient w

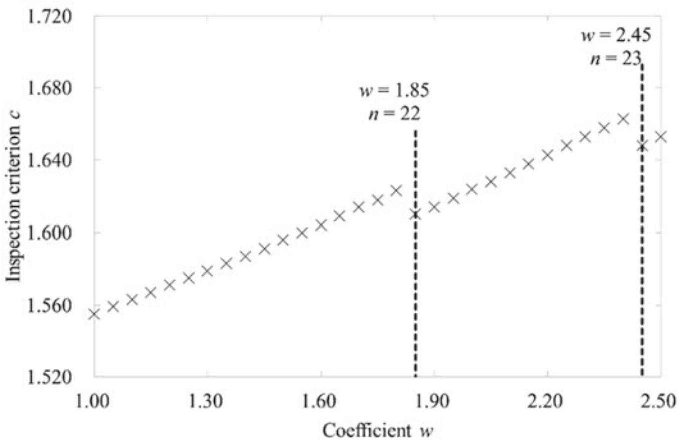


Fig. 2 Inspection criterion c by coefficient w

Then, the probability of non-acceptance has been verified. Table 1 shows the probability of non-acceptance when the product quality has been given as $\tau_1^2 = 2.5$. To verify the effect of bias, six pairs of (μ_1, σ_1^2) with the different mean and same quality loss have been computed. The probability has been computed by the Monte Carlo method using the normal distribution $N(\mu_1, \sigma_1^2)$. This simulation is computed $100,000$ times for every pair of (μ_1, σ_1^2). According to Table 1, the probability of non-acceptance, that is, the power shows all more than 0.90. Therefore, the proposed inspection satisfies the prescribed consumer's risk. When the mean is $\mu_1 \neq \mu_0 = 0.0$, the probability of acceptance increases according to the coefficient w. As we intended, the proposed inspection plan is sensitive to the change of the mean in the product quality. Figure 3 shows operating characteristic curve for difference in the coefficient w. Even if the sample size is same, the larger the coefficient w is, the smaller the probability of acceptance is when the mean in the product quality is different from the target value.

The proposed inspection can use any value of coefficient w. However, these values are not all desirable. When the value of w is large, the inspection sensitively detects the bias. From this perspective, a larger coefficient is desirable. On one hand, when the value of w is too large, the sample size will be increased. Since more sample size increases the cost, the same sample size as the case of $w = 1.0$ is needed. From this perspective, a too large coefficient is not desirable. Given these considerations, the desired coefficient w is defined as the maximum value when the required sample size is equal to the case of $w = 1.0$.

Table 1 Probability of non-acceptance in the case of $\tau_1^2 = 2.5$

w	(μ_1, σ_1^2)					
	(0.00, 2.50)	(0.55, 2.20)	(0.77, 1.90)	(0.94, 1.60)	(1.10, 1.30)	(1.22, 1.00)
1.0	0.909	0.908	0.913	0.922	0.935	0.952
1.1	0.909	0.911	0.920	0.932	0.947	0.965
1.2	0.909	0.913	0.925	0.940	0.957	0.975
1.3	0.908	0.915	0.930	0.947	0.964	0.981
1.4	0.907	0.917	0.934	0.952	0.970	0.985
1.5	0.907	0.918	0.937	0.956	0.974	0.988
1.6	0.906	0.920	0.940	0.960	0.977	0.991
1.7	0.905	0.921	0.942	0.963	0.980	0.993
1.8	0.904	0.922	0.944	0.965	0.982	0.994

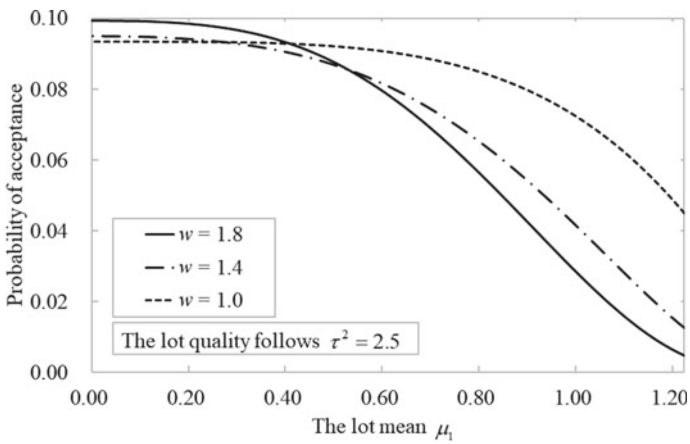


Fig. 3 Difference of operating characteristic curve by mean μ_1

5 Conclusion

This article has proposed the variable sampling inspection with emphasizing the difference between the mean in the product quality and the target value. For this purpose, inspection statistic with the coefficient was defined. Some numerical examples were calculated to confirm the operating characteristics of the inspection. The result has showed that the inspection can detect the bias sensitively. And in the case of using desired coefficient, it can be done while maintaining the inspection cost.

Acknowledgment This work was supported by JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2128.

References

1. Taguchi, G.: A Tutorial on Quality Control and Assurance—The Taguchi's Methods. Annual Meeting, Las Vegas, Nevada (1985)
2. Arizono, I., Kanagawa, A., Ohta, H., Watakabe, K., Tateishi, K.: Variable sampling plans for normal distribution indexed by Taguchi's loss function. Nav. Res. Logist. **44**(6), 591–603 (1997). https://doi.org/10.1002/(SICI)1520-6750(199709)44:6%3c591::AID-NAV5%3e3.0.CO;2-Z
3. Yen, C.-H., Chang, C.-H.: Designing variables sampling plans with process loss consideration. Commun. Stat. Simul. Comput. **38**(8), 1579–1591 (2009). https://doi.org/10.1080/03610910903046809
4. Aslam, M., Yen, C.-H., Chang, C.-H., Jun, C.-H.: Multiple state repetitive group sampling plans with process loss consideration. Appl. Math. Model. **37**(20–21), 9063–9075 (2013). https://doi.org/10.1016/j.apm.2013.04.020
5. Wu, C.-W., Wang, Z.-H.: Developing a variable multiple dependent state sampling plan with simultaneous consideration of process yield and quality loss. Int. J. Prod. Res. **55**(8), 2351–2364 (2016). https://doi.org/10.1080/00207543.2016.1244360
6. Patnaik, P.B.: The non-central χ^2- and F-distributions and their applications. Biometrica **36**, 202–232 (1949). https://doi.org/10.1093/biomet/36.1-2.202

Proposal and Verification of Quality Control Method by Adjusting Process Mean in Post-process

Akimasa Otsuka, Kazuya Hayashi, and Fusaomi Nagata

Abstract Variations of component quality have a significant impact on product quality and its performance. The variations in the manufacturing process are due to variations of dimensions and geometrical features caused by machining errors. The selective assembly method is a standard method for improving product quality. However, it takes time and costs much because all products should be measured, classified in rank, assembled into a product. In this study, we proposed a method to stabilize product quality by managing the parts that are produced machined in a later process. The effectiveness of the proposed method is verified in simulations using a virtual product consisting of two parts. Following two cases were simulated to verify the effectiveness, Case A where the process mean is randomly set assuming a situation where two parts are made individually, and Case B where the process means of the part machined later is controlled according to process means of another part machined before. The cases A and B are compared by calculating values of process capability index Cpk calculated by the process mean and standard deviation when the two parts are assembled. In order to visualize the fluctuations of the process mean and process capability index, control charts of Cpk for each lot and investigated the Cpk statistics. The results showed that the proposed method could significantly decrease a variation of process mean and implied that specifying limits of process capability index in design drawing can realize quality design in mass production.

Keywords Process control · Process capability index · Quality design · Mass production

1 Introduction

Recently, process control in the manufacturing process is evolved because elements of the manufacturing process, such as sensors, robots, software, the internet, and artificial intelligence, are also evolved day by day. By exploiting those components

A. Otsuka (✉) · K. Hayashi · F. Nagata
Sanyo-Onoda City University, Yamaguchi, Japan
e-mail: otsuka_a@rs.socu.ac.jp

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_74

in the manufacturing process, future products will have higher quality, and manufacturing costs will decrease. This future situation will be welcomed around and can be achieved by developing informatics concepts and techniques. Integrating modern techniques, Industrial Internet of Things (IIoT), Artificial intelligence, Digital twin, and big data, the design and manufacturing processes of a product may significantly change [1].

Generally, a product consists of several assemblies. An assembly usually consists of several parts. The quality of the product depends on the qualities of the parts, and the process quality of assembling the parts is also important. In the manufacturing process of parts, variations of dimensions and geometrical features of parts are caused by several machining errors due to fixing-error of materials, kinetic-error of machine tools, and assembling-error. If those errors were zero, the high-quality assembly could be easily realized. Up to now, many studies have been addressed to control the quality effectively.

The selective assembly method is a common and traditional method for achieving high-quality. The selective assembly is applied for products that require high precious assembling, such as mating parts of slide bearing, ball bearing, etc. Selective assembly is still studied to realize a higher quality of assemblies [2]. However, it takes time and costs much because all products should be measured, classified in rank, and assembled into a product.

In this paper, a method in terms of process control is proposed. The method is controlling the size of a part according to the size of the other part machined in advance. For example, products usually have mating parts consisting of a pin and a hole, and clearance is essential for product quality. The clearance is the difference between the pin size and the hole size. In mass production, those two parts are usually produced individually then assembled each other. If we control the process of the hole parts according to the mean size of the pin parts produced in advance, the mean size of the clearance will be reduced.

The method needs an assumption that the process mean of the hole parts is observable and follows the standard distribution. As the method may need to produce them sequentially, it may take much time to produce the assemblies. However, the method could stabilize product quality by managing the parts that are produced machined in a later process. The effectiveness of the proposed method is verified through simulations using a virtual product consisting of two parts. The result showed that the proposed method could significantly decrease a variation of process mean. The simulation results imply that specifying limits of process capability index in design drawing could realize quality design in mass production.

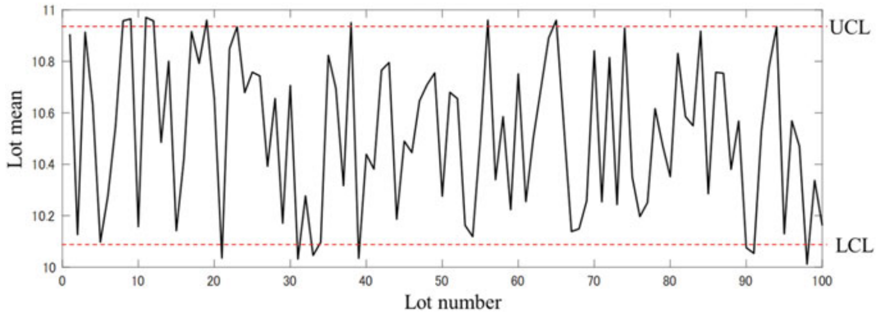


Fig. 1 Example of X-bar chart

2 Process Control

2.1 Control Chart

Control charts are a traditional method of statistical quality control and are still a typical tool. Control charts are used to visualize process stability, products pass or fail judgments, and process adjustments. A stable state of the process must be achieved and maintained for producing high-quality outputs. The basic function of the control chart is to statistically visualize variation of process mean caused by special causes as opposed to process variance caused by common causes. Common causes change inherent in the process in the long run. If the variation parameters are kept between the upper control limit line and the lower control limit line, the process is in a stable state. A product consists of several parts, and information on the control charts is feedbacked to control the part's quality in the machining process.

The most standard control chart is the Shewhart control chart that has upper and lower control limits. The control limits are decided based on tolerance and process standard deviation. Figure 1 shows an example of an X-bar chart, in which horizontal and vertical lines are respectively lot number and lot mean (lot mean). If the lot mean is larger than UCL or smaller than LCL, the lot is rejected. Products quality is assured lot by lot by observing this chart, and the defects of the process can be found live.

2.2 Assumption of the Normal Distribution

A random error of parts size statistically follows the normal distribution. If those parts are assembled, the assembly size also follows the normal distribution. Consequently, the mean of the distribution follows the normal distribution. In mass production, the assemblies are produced and inspected lot by lot. The lot is controlled by monitoring its range, mean, standard deviation. Under process control, the process mean is usually assumed to be the normal distribution, and this paper also assumed it. The standard deviation of the process follows the chi-square distribution if the process follows the normal distribution.

2.3 Process Capability Index

Process capability index Cpk is an index for evaluating the capability of a manufacturing process from the process mean and standard deviation [3]. The process capability index is a dimensionless index that indicates how well the quality of the manufacturing process is controlled. There are several types of indices, such as Cp, Cpk, Cc, and Cpm. Process capability index Cpk is commonly used because Cpk is related to the defects rate of products. Definition of Cpk is shown as the following equation.

$$\mathrm{C_{pk}} = \min\{(\mu - L)/3\sigma, (U - \mu)/3\sigma\} \quad (1)$$

where μ is the population mean, σ is the population standard deviation, L is the lower tolerance limit, U is the upper tolerance limit. In this paper, process capability index Cpk is used to quantify lot by lot qualities of mass-produced products. Cpk can be used when the process is stable and under statistical process control.

As general criteria for process capability index, if Cpk is lower than 1.00, then the process is regarded as low capability. If Cpk is 1.33, then the process is regarded as normal capability. If Cpk is larger than 1.33, then the process is regarded as high capability. Process capability control charts have been studied [4], and it is still a useful tool. In this study, process capability index Cpk is evaluated to quantify product quality in simulation.

3 Method for Controlling the Quality of Products Lot by Lot

In this study, we proposed a method to stabilize the product quality by managing the parts machined in the latter process. Figure 2 shows an overview of the proposed method for a product consisting of two parts. The product specification is virtual, for example. The method requires assumptions that part 1 and part 2 are machined separately, and the part 1 is machined after machining and measuring another lot. Under those assumptions, the target size of part 2 machined later is controlled by the mean size of part 1. Assembling two parts into a product, the mean size of products the lot can be controlled better. The effectiveness of the proposed method is verified in simulations using a virtual product consisting of two parts.

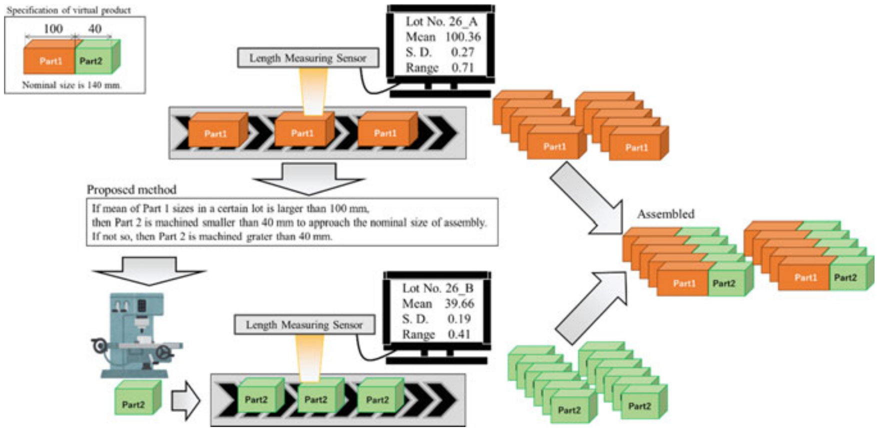


Fig. 2 Overview of the proposed method

4 Case Study in Simulation

4.1 Case Study

Following two cases were simulated to verify the effectiveness, Case A where the process mean is randomly set assuming a situation where two parts are made individually, and Case B where the process mean of the part machined later is controlled according to the process mean of another part machined before. Case A and Case B are compared by calculating values of process capability index Cpk calculated by the process mean and standard deviation when the two parts are assembled. In order to visualize the fluctuations of the process mean and process capability index, control charts of Cpk for each lot and investigated the Cpk statistics. The result showed that the proposed method can significantly decrease a variation of process mean. This result implies that specifying limits of process capability index in design drawing could realize quality design in mass production.

4.2 Simulation Conditions

A simulation with different conditions is performed to judge the result of this proposal. Case A is the case where part 2 is manufactured regardless of the size of part 1. This is called the traditional method. Case B is the case where the size of part 2 is changed according to the size of part 1. This is the proposed method. The variation of the process means is within the range of the standard. The variability is randomly selected from the normal distribution, and the variability is added to the previous process mean to obtain the next process mean so that it is influenced by the product manufactured before. Standard deviations of both part sizes are fixed in this study. The number of lots is set to 1000, the number of products in a lot is 50 pieces, the size of Part 1 is 100.0, tolerance of Part1 is 0.14, the size of Parts 2 is 40.0, and tolerance of Part 2 is 0.10. Tolerance is set to be complying with ISO IT10 grade.

5 Results

To compare each product quality of Case A and Case B quantitatively, Cpk values are calculated by Eq. (1) for every assembly lot. Figure 3 shows Cpk values of whole lots after the simulation. A horizontal line is the lot number, and a vertical line is the Cpk value of the lot. Comparing the result of the traditional method and proposed method, the Cpk variance of the proposed method is significantly smaller than Case A. Figure 4 is an enlarged view of Fig. 3 and is shown from lot 1 to lot 50 to show the smaller variance of Cpk values of the proposed method.

The means of the Cpk values of Case A and Case B through 1000 lot simulations are both 1.00. The standard deviations of the Cpk values of them are 0.0168 and 0.0108, respectively. There is a difference about 1.6 times in standard deviation. It is also found that the minimum value of the Cpk of Case B is about 0.1 higher than the value of Case A. It is concluded that the products produced by the proposed method have higher quality in terms of less variation of products.

However, there is a tradeoff between smaller variance and longer production time in the proposed method because part 2 should be produced after part1 is produced and measured. Therefore, the method could be efficiently applied if the order of parts productions is fixed.

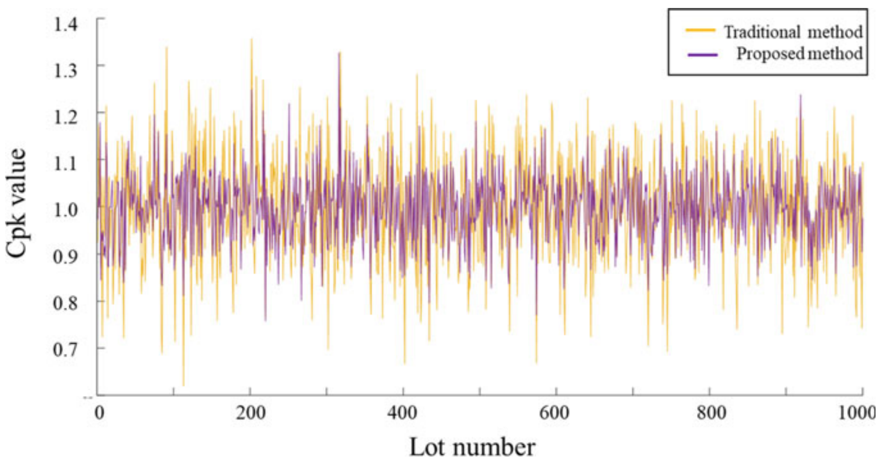


Fig. 3 Simulation results from 1 to 1000 lot numbers

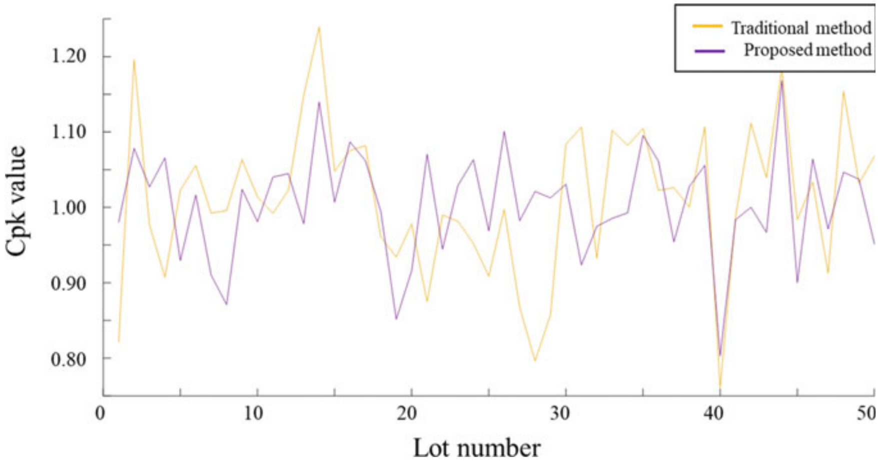


Fig. 4 Simulation results from 1 to 50 lot numbers

6 Conclusion

In this paper, a method to stabilize product quality by managing the parts produced machined in a later process was proposed. The effectiveness of the proposed method was confirmed in simulations using a virtual product consisting of two parts. The result showed that the proposed method could significantly decrease a variation of Cpk. This result implies that specifying limits of process capability index in design drawing could realize quality design in mass production. For future work, the method will be expanded for a real product consisting of several parts. Furthermore, this proposal should be compared the existing method called selective assembly.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number 20K04225.

References

1. Sanz, E., Blesa, J., Puig. V.: BiDrac Industry 4.0 framework: application to an automotive paint shop process. Control Eng Pract **109**, 104757 (2021). https://doi.org/10.1016/j.conengprac.2021.104757
2. Aderiani, A.R., Warmefjord, K., Soderberg, R., Lindkvist, L.: Developing a selective assembly technique for sheet metal assemblies. Int. J. Prod. Res. **57**(22), 7174–7188 (2019)
3. Kane, A.V.: Process capability indices. J. Qual. Technol. **18**(1), 41–52 (1986)
4. Chatterjee, M.: Exact expression for the expectation of estimated Cpk based on control chart information and the corresponding process capability control charts. J. Stat. Comput. Simul. **87**(5), 1025–1047 (2017)

Human Factors Engineering

Examining the Correspondence of Cognitive Costs and the Mutual Information Criterion in Rational Inattention Models

Qi Wu, Shinji Nakazato, Bojian Yang, and Tetsuya Shimokawa

Abstract The rational inattention model has recently attracted much attention as a promising candidate to model bounded rationality in the research field of decision-making and game theory. The model presumes that the cost of information processing is proportional to the mutual information obtained from signals. It has been reported that the introduction of this information cost can explain various phenomena observed in the market, and applied in a wide range of fields such as finance, bargaining, auction, and policy analysis. However, the rational inattention model does not have a sufficient cognitive foundation, despite the amount of attention it has received. In this study, we will use both a behavioral and neural approach to investigate whether the amount of mutual information corresponds to cognitive costs.

Keywords Rational inattention · Mutual information · Experiment

1 Introduction

Theoretical models that presume human rationality are the mainstream in the field of finance and game research, but in recent years, the rational inattention (RI) model, which explicitly introduces cognitive costs associated with information acquisition and processing, has been attracting attention. The RI model is characterized by the assumption that the cognitive processing cost of information can be measured in terms of the amount of mutual information from the acquired signal. Once this assumption is accepted, theoretically, the existing model can be extended to a more realistic decision-making setting in a relatively natural way, and it also has the great advantage of being able to use the accumulated knowledge in the field of information

Q. Wu (✉) · S. Nakazato · B. Yang · T. Shimokawa
Tokyo University of Science, Tokyo, Japan
e-mail: shkidwq@hotmail.com

B. Yang
Sony Group Corporation, Nihonbashi, Japan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_75

theory. The RI model has been actively studied as a theory that links decision-making and information theory, and as a modern theory that expresses Simon's bounded rationality [1–3].

The history of the RI models begins with Sims [4, 5], and there have been two major streams of research. The first is the Kalman filter type model initiated by Sims [4]. This model is used in a dynamic environment, in which the accuracy of the filter, or Kalman gain, is determined by the cognitive capacity constraints of information processing. This cognitive capacity, like that in Shannon's sense, is an upper limit on the total amount of information that can be processed, but in Sims [4] it is determined by the subject's ability. This model has been applied to many dynamic models, including financial and policy analysis.

Another stream is the stochastic choice type model. This model assumes that information processing costs are incurred in proportion to the amount of mutual information, and given this assumption, the subject decides not only the action to take but also the amount of information to use to maximize the expected utility. This model is known to be closely related to logit-type stochastic choice models [6] and has been applied to a variety of fields including game theory [7]. Over recent years, there has been progressing in the refinement of information cost modeling and in extending to dynamic environments.

However, in contrast to this energetic promotion of the theoretical works, empirical verification of the validity of the RI model has not progressed much. There have been only a few behavioral experiments and even fewer in more realistic settings. Furthermore, to our knowledge, the central assumption of the RI model, that the amount of mutual information obtained from signals adequately represents the cognitive cost of information, has not been tested from a neuroscientific perspective. The present study aims to test this point.

Specifically, we will examine whether the cognitive processing cost of the model is consistent with the activation of brain regions related to working memory and reasoning, such as the dorsolateral prefrontal cortex, ventral prefrontal cortex, and rostral prefrontal cortex. The difficulty in empirically testing the RI model is that we cannot directly measure the amount of mutual information, because this amount is subjectively determined depending on the amount of signal used by the subject, and thus is not available to the observer. In this paper, we avoid this problem by focusing instead on the parameters representing the capacity in the Kalman filter-type model and the information cost in the stochastic choice type model.

The contributions of this paper are twofold:

1. We first tested the RI model from a neuroscientific point of view, and validated its central assumption that the cognitive processing cost of information can be measured in terms of the amount of mutual information from the acquired information, by mapping it to the activation of the prefrontal cortex, which is well known in cognitive science as costly cognition.
2. We proposed an experimental framework, a sequential investment task, in which the two main models of RI can be treated simultaneously in a more realistic experimental environment.

2 Problem Setting

2.1 Sequential Investment Task

In this task, participants make predictions about the return of a price sequence each period and decide whether to invest in a single stock or a safe asset. The participant's goal is to maximize the expected return. The fundamental return of the stock is randomly set at the beginning of each price sequence. The safe asset return is always set to zero. This fundamental return is fixed and unchanging throughout the price sequence, but the actual return of the stock observed each period is composed of the fundamental return plus white noise caused by exogenous factors such as the market environment. Therefore, participants cannot directly observe the fundamental returns when making investment decisions. However, they can calculate a sufficiently large number of past returns from observable price sequences and average them to get the fundamental return. If the inferred fundamental return is higher than the safe asset return, the rational decision in this setting is to invest fully in the stock, otherwise not.

In this experiment, we have two treatments for the variance of exogenous noise, one large and one small. Each price column was prepared with 10 paths with different fundamental returns. Five of these paths used white noise with large variance, and the other five paths used white noise with small variance.

Let f be the return determined from the fundamentals of the stock and be constant over time. The return observed in each period t is defined by $r_t = f + \varepsilon_t$, where $\varepsilon_t N(0, \sigma_N^2)$ $i.i.d.$. The return on the safe asset is always assumed to be $r^S = 0$. The price sequence presented to the participant's GUI is calculated as $P_{t+1} = (1 + r_t)P_t$. Therefore, when the fundamental return is zero, $f = 0$, the price will randomly walk.

The participant can predict the fundamental return f from the observed returns. If there is no cognitive constraint on information processing, fundamental returns f can be identified from the law of large numbers by using a large enough number of past returns and taking the average of the time series. However, if there is a cognitive constraint, participants may not reasonably formulate a prediction because it takes a certain amount of calculation to obtain each period's returns from prices and to take their average.

To fit the setting of the model to be analyzed, the decision-making of the participants was assumed to be static. That is, participants are given 100 units of points at the beginning of each period, which is cleared at the end of the period. Thus, the participant decides each period whether to invest in a stock or safe asset to maximize the return of the period. However, participants can use the historical price sequence to identify the fundamental return in this experiment. Participants are first presented with a 100 periods price sequence. The reason for presenting 100 periods price in advance is to give a sufficient sample to infer the fundamental return. Participants then invest, the next price sequence is revealed, and liquidation for that period takes place. Price updates are made every second. Participants made this independent choice 180 times per price sequence.

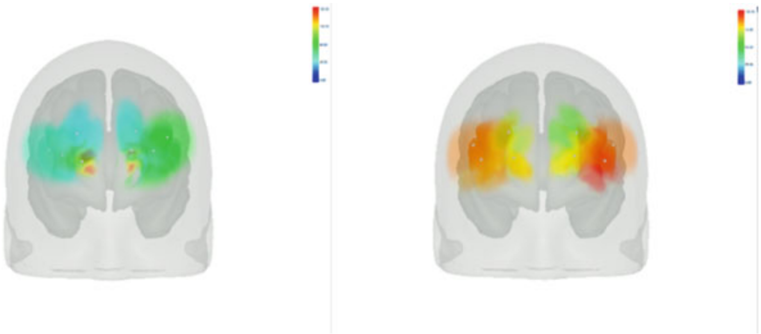


Fig. 1 NIRS image

2.2 Experimental Setup

The biometric information used in this analysis was the change in blood hemoglobin concentration in the prefrontal area. Functional NIRS (BriteMKII supplied by Artinis Medical Systems) was used to measure the blood hemoglobin concentration in the prefrontal area. Figure 1 shows a visualization of the change in blood hemoglobin concentration during the experiment. The left figure shows the activation of the ventral region, and the right figure shows the activation of the dorsolateral and dorsoventral regions. We will focus on the dorsolateral, ventral, and rostral regions, which are considered to be closely related to costly cognition, working memory, and reasoning [8–11].

Functional NIRS measures the difference in optical absorption characteristics between oxygenated and non-oxygenated hemoglobin by irradiating near-infrared region light from an external source. Therefore, compared to fMRI, it is more difficult to measure deep brain regions, but on the other hand, it has the advantage of being able to measure in a non-constrained and realistic environment. The sampling rate is relatively fast (50 Hz), but the spatial resolution is not so high (20–30 mm). This is non-invasive equipment. These characteristics make it suitable for decision analysis in more realistic environments as in the present financial task.

2.3 Participants

The participants of this experiment were 40 undergraduate and graduate students of the School of Business, Tokyo University of Science. None of them had any investment experience. The number of valid samples was 39. Subjects made 180 investment decisions for one price sequence and were rewarded up to 1000 yen (including 500 yen for participation) according to their points. Subjects were randomly assigned to a task for each treatment. 19 subjects were assigned to the large white noise condition and 20 to the small condition. Subjects were briefed on the GUI operation and the investment task setting and then engaged in the task.

3 Analysis Model

3.1 The Kalman Filter Type RI Model

In the RI model by Sims [4], the subject attempt to remove the noise in the observed signal (past returns in the present task) by using the Kalman filter, but the accuracy of the filter is constrained by the subject's information processing capacity. If the capacity is small, the information in the signal cannot be utilized well (Kalman gain becomes small), which means that there is a momentum in investment.

More specifically, it is modeled as follows. We write the subjective variance of the noise in the observed signal as σ_N^2. The conditional mutual information of the observed return r_t in each period t is $I_{t-1}(f) = \frac{1}{2}\log(1 + \sigma_N^{-2}\sigma_{t-1}^2)$, and the constraint $\kappa = \frac{1}{2}\log(1 + \sigma_N^{-2}\sigma_{t-1}^2)$ holds when the cognition capacity of a subject is κ. From the constraint, the observed noise becomes $\sigma_N^2 = \frac{\sigma_{t-1}^2}{\exp(2\kappa)-1}$. In this model, the noise appearing in the observation equation is therefore subjective; if κ is ∞, then the variance of the observation noise is 0, and if κ is 0, then the variance of the observation noise is ∞. Given the subjective observation noise, the update of people's expectation follows;

$$\mu_t = \mu_{t-1} + \frac{1}{1 + \sigma_N^2(\kappa)}(r_{t-1} - \mu_{t-1}),$$
$$\sigma_t^2 = \sigma_{t-1}^2 - \left(\frac{\sigma_{t-1}^2}{\sigma_{t-1}^2 + \sigma_N^2(\kappa)}\right)\sigma_{t-1}^2 \tag{1}$$

People invest in the stock or safe asset based on the expectation formation. In the following analysis, we examine six variations of the choice model, including loss aversion, are examined. The details of the models are described in Sect. 4.2.

3.2 Stochastic Choice Type RI Model

In the stochastic choice type RI model, the subject is assumed to make a decision in two steps [6]. The first step is to choose the information strategy to use, and the second step is to choose the action to optimize the expected profit based on the prediction/belief of the fundamental return. It is optimal to use as many and as useful signals as possible to identify the fundamental returns, but a certain percentage of cognitive cost λ is required in proportion to the amount of mutual information obtained from the signals.

More specifically, in the second stage, the subject decides whether to invest in the stock or safe asset according to the conditional distribution of the fundamental return $p(f|s)$ given signal s. The utility realized with an optimal action at this stage is denoted by $V(p(f|s))$. In the first stage of information strategy selection, given the expected utility, the subject decides which signal s to use. The choice of information strategy determines which signal structure $p(s, f)$ is desirable. To obtain more detailed information, a cost $\lambda\{H(f) - H(f|s)\}$ is incurred according to the amount of mutual information, where H is the Shannon entropy, and the amount of mutual information is defined as the decrease in entropy due to the acquisition of the signal. λ is a parameter that represents the cognitive cost per unit of mutual information. The first-stage problem can be written as follows:

$$\max_{p(s|f)} \sum_{f} \sum_{s} V(p(f|s))p(s|f)p(f) - \lambda\{H(f) - H(f|s)\},$$

where we assume that the possible states of the fundamental return are finite to keep the discussion simple.

Assuming that the ex-ante choice probabilities of a stock and safe asset are equal before the fundamental returns are given in the experiment (the expected value of the fundamental return f given in this experiment is zero, which is equal to the return of the safe asset), the solution to this problem can be derived by the following softmax-type choice rule.

$$p_t = \exp\left(\frac{f}{\lambda}\right) \Big/ \left\{1 + \exp\left(\frac{f}{\lambda}\right)\right\}, \quad \forall t$$

To derive this rule from the above optimization problem, some equation transformations are required. See Lemma 1 and Collorary 1 in Matějka and McKay [6] for more details on this.

4 Empirical Results

4.1 Quantile Analysis

We examine the relationship between the parameter values of each RI model and the hemoglobin concentration in the blood. As mentioned earlier, the amount of mutual information obtained from signals is determined subjectively, and this cannot be observed directly. The parameters of each model can be considered as its alternative indicators. Figures 2 and 3 show the average hemoglobin concentration in blood of each group in dichotomous analysis, and Fig. 4 shows the average hemoglobin concentration in blood of each group in the trichotomous analysis. The figure on the left shows the hemoglobin concentration in the blood of each sample, divided into two or three groups according to the size of the estimated model parameters (the left bar shows a small value group and the right bar is a large value group). From left to right, they correspond to rostral, dorsal-extralateral, and ventral-extralateral regions. The right figure is a scatter plot of the parameters for each sample and the blood hemoglobin concentration in the extraperitoneal region.

From these figures, we can see that brain activity is consistent with the assumptions of the RI models. In other words, the larger the cognitive cost λ and the larger the capacity κ, the more activated the brain regions involved in costly cognition. The differences between these groups are significant for λ in all brain regions in the dichotomous partition and all brain regions except the rostral region in the tripartite partition (5% level in ANOVA analysis). However, for κ, no significant difference was confirmed for any of the regions. As can be seen from the scatter plots in each figure, the correlations are not sufficiently clear due to the large individual differences in biological data. This may be partly because the Kalman filter type RI model does not fit well in the present sample, which will be revealed in the next section.

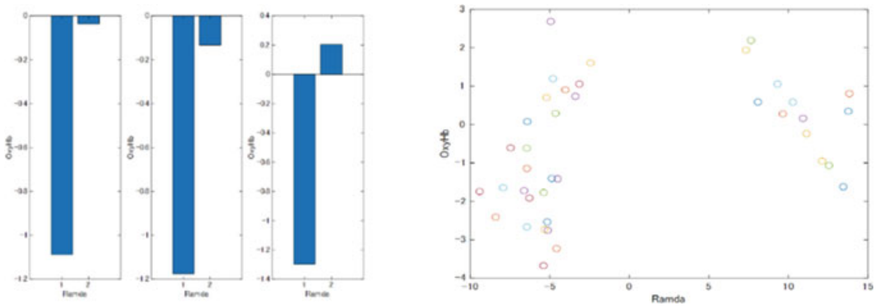


Fig. 2 Correlation with λ, dichotomous analysis

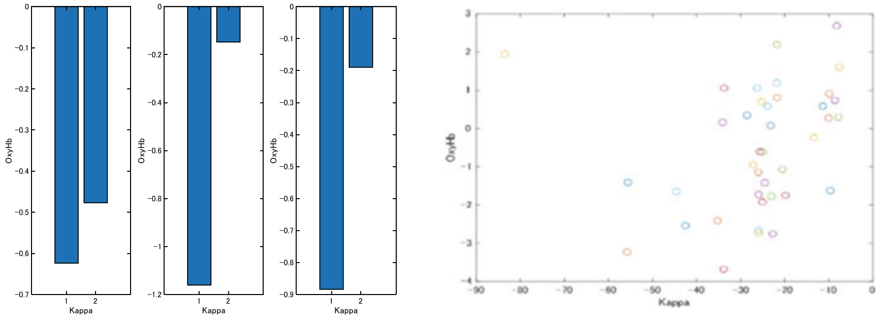


Fig. 3 Correlation with κ, dichotomous analysis

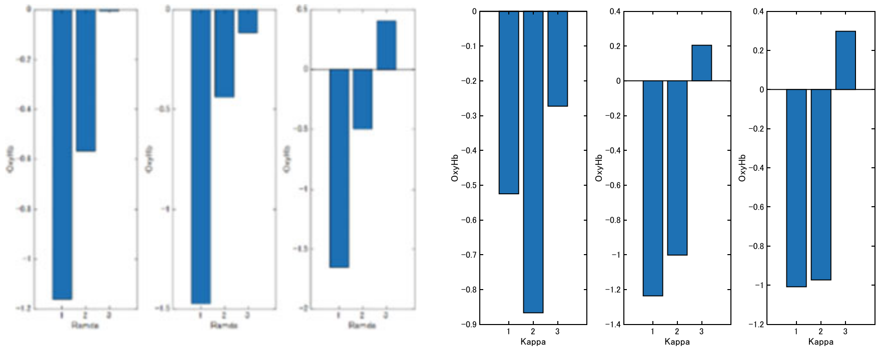


Fig. 4 Trichotomous analysis, λ on the left and κ on the right

4.2 Model Selection

Finally, we selected a more appropriate model based on the behavioral data, using the maximum likelihood method for model fitting. The table (Model selection) summarizes the log-likelihood (llh), Akaike's information criterion (AIC), and R^2 (R2) for the Kalman filter type and stochastic choice type RI models and their variations. R2 is the improvement in the log-likelihood from a random prediction, defined as follows, $R^2 = \frac{\log M - \log M_{rand}}{-\log M_{rand}}$, where M is the likelihood of the model concerned and M_{rand} is that of the random prediction model.

We examined the following six variants of the Kalman filter type RI models and the stochastic choice type RI model. Table 1 clearly shows that the stochastic choice type is superior for this experimental data. Since the fundamental returns were fixed in this experiment, the prediction converged at a relatively early stage, and therefore the characteristics of the Kalman filter type model may not have been fully utilized. As for the change in blood hemoglobin concentration, the correlation between the parameters of the stochastic choice type model was clearer than that of the Kalman filter type model, the same tendency was observed for the behavioral data.

Table 1 Model selection

	Stochastic choice type	Kalman filter type					
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
llh (mean)	− 106.8063496	− 123.925	− 114.441	− 118.459	− 122.949	− 123.221	− 112.993
llh (std)	28.19307684	0.619818	20.88867	13.03432	1.804096	1.17822	23.27223
AIC	215.6126993	249.8496	230.8825	240.9178	249.8971	250.4413	229.9866
R2	0.139167648	0.001197	0.077632	0.045251	0.009066	0.006873	0.089302

References

1. Simon, H.A.: A behavioral model of rational choice. Q. J. Econ. **69**(1), 99–118 (1955). https://doi.org/10.2307/1884852
2. Simon, H.A., Barenfeld, M.: Information-processing analysis of perceptual processes in problem solving. Psychol. Rev. **76**(5), 473 (1969). https://doi.org/10.1037/h0028154
3. Mackowiak, B., Matejka, F., Wiederholt, M.: Rational Inattention: A Review. CEPR Discussion Papers 15408 (2020)
4. Christopher, A.: Sims: Implications of rational inattention. J. Monet. Econ. **50**(3), 665–690 (2003). https://doi.org/10.1016/S0304-3932(03)00029-1
5. Christopher, A.: Sims: rational inattention: beyond the linear-quadratic case. Am. Econ. Rev. **96**(2), 158–163 (2006). https://doi.org/10.1257/000282806777212431
6. Matejka, F., McKay, A.: Rational inattention to discrete choices: a new foundation for the multinomial logit model. Am. Econ. Rev. **105**(1), 272–298 (2015). https://doi.org/10.1257/aer.20130047
7. Matejka, F.: Rigid pricing and rationally inattentive consumer. J. Econ. Theor. **158**, 656–678 (2015). https://doi.org/10.1016/j.jet.2015.01.021
8. D'Esposito, M., Detre, J.A., Alsop, D.C., et al.: The neural basis of the central executive system of working memory. Nature **378**(6554), 279–281 (1995). https://doi.org/10.1038/378279a0
9. Smith, E.E., Jonides, J.: Neuroimaging analyses of human working memory. Proc. Natl. Acad. Sci. **95**(20), 12061–12068 (1998). https://doi.org/10.1073/pnas.95.20.12061
10. Smith, E.E., Jonides, J.: Working memory: a view from neuroimaging. Cognit. Psychol. **33**(1), 5–42 (1997). https://doi.org/10.1006/cogp.1997.0658
11. Osaka, M., Osaka, N., Kondo, H., et al.: The neural basis of individual differences in working memory capacity. Neuroimage **18**(3), 789–797 (2003). https://doi.org/10.1016/S1053-8119(02)00032-0

A Proposed Model for 3D Printing Education in the University

Anca Draghici, Carmen Sticlaru, Agneta Lovasz, and Larisa Ivascu

Abstract 3D printers are a fascinating tool for engineering teaching and education, related to a large variety of subjects. More and more universities are integrating not just the topic of additive manufacturing, but 3D printers' infrastructures to create great learning experiences. The main scope of the paper is to analyze the state of teaching and learning 3D printing in the universities, presenting the advantages, particularities, and the involved actors. There will be supported the ideas that through 3D printing, students can translate their ideas directly into reality, and spatial imagination. Usually, initial teaching and learnings activities begins with simple physical objects and later deals with abstract, virtual 3D models and complex assemblies. The "magic" of teaching and learning 3D printing is that it allows quick reversal, from the 3D CAD drawing to the physical object; the direct link of the two processes is stimulating creativity and enhance imagination. Finally, there will be discussed the case of teaching and learning 3D printing at Politehnica University of Timisoara (Romania) with the support offer by the "3D Printing Support Service for Innovative Citizens" INNO3D project (2019-1-IE203-000693INNO3D).

Keywords Education · 3D printing · Training program · Education model · INNO3D project

A. Draghici (✉) · C. Sticlaru · A. Lovasz · L. Ivascu
Politehnica University of Timisoara, Timisoara, Romania
e-mail: anca.draghici@upt.ro

C. Sticlaru
e-mail: carmen.sticlaru@upt.ro

A. Lovasz
e-mail: agneta.lovasz@upt.ro

L. Ivascu
e-mail: larisa.ivascu@upt.ro

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_76

1 Introduction

The engineering practices of today required new interdisciplinary skills and competencies (transversal, trans-sectoral), which require the ability of a person to connect and integrate knowledge from different fields couple them to a specific context; this approach involved a lot of creativity, agility, and innovation. This is engineers' actual ways of acting in design, manufacturing, logistics, data analysis etc. These competences can be developed through an effective physics instruction, which would be able to guide the students towards a deeper understanding of disciplinary fundamental concepts and, at the same time, to strengthen their reasoning abilities and transversal skills" [1, 2]. Furthermore, recent updates of the pedagogy procedures, standards of science education strongly encourage the design and use of instructional environments focused on the engagement in the engineering practices, as a more efficient and effectiveness way of teaching and learning students enrolled in engineering programs [1, 2]. At this regard, the use of 3D printing technology, placed in a problem/project-based learning approach, to foster engineering education has been proved as a feasible solution. It has been reported the improvement of engineering design teaching and learning process based using 3D printing and that students could rapidly and directly measure the created artefact quality [3]. Teaching staff experiences from different university programs are convergent on recognized that efficient of the comprehensive abilities development by using 3D printing technology within curricula subjects related to the engineering studies [3–9]. The extended use of 3D printing in higher education is related with STEM (Science, Technology, Engineering and Mathematics) education development and the modern approach of training the new generation of future engineers. This, FabLabs or MakerSpaces of universities were created using international projects or public–private partnership [10–12].

In this context, the present article will present a brief literature review referring to teaching and learning 3D printing in universities, presenting the advantages, particularities, and the involved actors. There will be supported the ideas that through 3D printing, students can easy transfer their ideas into tangible products, and spatial imagination. The literature review will be accompanying by facts on industrial use of the technology as motivation factors for students to achieve the related competences. Second, there will be discussed the case of teaching and learning 3D printing at Politehnica University of Timisoara (Romania) with the support offer by the INNO3D project.

2 Literature Review and Industrial Development

2.1 Literature Review

In 2014, a research has reported that 3D printing could directly support the teaching process [13]. From the idea to the printed part, there have been described an easy process to follow in the university laboratory, which strongly motivates students, and promotes understanding of the engineering design process and enhances students' involvement in class; by using 3D printing technologies, students can translate their ideas directly into reality, and spatial imagination is promoted [13]. Usually, in the early years, teaching and learning engineering design begins with physical objects and later deals with abstract, virtual 3D models. But introducing the innovative 3D printing technologies allows a reversal: from drawing (3D CAD) to the physical object. It is now possible to directly link the two processes during the learning phase, stimulating creativity and enhance imagination [3–9].

The use of 3D printing in engineering education and training were growing rapidly. All around the world, higher education institutions are developing new curricula and purchasing equipment, software, and materials to support research and instruction. In response to the growing opportunities in the additive manufacturing job market, the interests among students in this field of study have increased dramatically [3–9]. In addition, the engineering education in this field has been supported by the tremendous equipment developers and providers, being recognized that the 3D printers' market that has rapidly evolve [14]; actually, it is very easy to buy and to use a desktop 3D printer at home and this is because the "increase demand for smaller, more affordable 3D printers that has grown amongst hobbyists, do-it-yourselfers, design engineers, and small business owners". There have been observed that desktop 3D printer providers and their market continuously expand as many existing and new companies release printers with greater capabilities and better features [14].

In addition, universities have supported the FabLabs or MakerSpaces development [10–12] as collaborative working places, with adequate digital and non-digital infrastructure (spaces with equipment such as: computers, printers, laser cutters, 3D printers, CNC machines, woodworking tools, soldering equipment, sewing machines etc.) and coaches (trainers, demonstration, addition learning sessions etc.) that contributed to the practical work in engineering projects; the aim is to provide non-specialists the access to technologies, so that they can explore, learn and make [10]. Using such techno-creative activities and spaces of universities [15], students could develop their technical skills and competences by practical exploration of the 3D printing technologies and thus, better connect theoretical knowledge with applications. Furthermore, in a recent published study [10], there have been recognized that FabLabs and MakerSpaces "spaces go beyond digital twenty-first century skills as they are mixed digital and physical environments". The learning solution has expanded internationally being recognized that FabLabs are those makerspaces that have signed the Fab Charter of the Fab Foundation; actually, more than 1,750 active FabLabs centers exists all over the world [16].

Trainers and learners have recognized that engineering education and training can be benefited through 3D printing by creating excitement in learning and educational practices, most by supporting STEM curriculum, opening new possibilities for learning, critical and analytical thinking, innovation, and entrepreneurship, giving access to new materials and applications not available before and promoting problem-solved skills and methodologies. Using the 3D printing technologies students can take their design ideas to the next level by allowing them to get involved in the fabrication stage [2–4, 6–8]. In many higher education institutions, 3D printing has been integrated into conventional undergraduate and graduate programs, offering advanced training programs or courses only [5–8, 17, 18]. At this level of education, important use is to integrate 3D printing technologies in sciences (e.g., the case of applying in different engineering subjects, programs). Models are produced to support students' knowledge acquisition in a practical manner [2, 5–9]: developed test models can be used for experiments and test specimens can be useful for learning about mechanical properties of materials. Teaching staff has recognized that educational approaches using 3D printing differ in duration being differentiated by the learning objectives, and the competencies that are to be taught. Numerous and diverse methods and tools, formats, which are used by universities to teach 3D printing technology refers to: integration into existing teaching units; provision of 3D printing literature support by university library; online tutorials (with certificates); Independent Bachelor or Master Lecture (e.g., Rapid Prototyping, Additive Manufacturing, 3D Printing); 3D printing/additive manufacturing course of studies (already included in the master programs); 3D Printing Summer Schools; Voluntary 3D printing working groups (dedicated to a specific research or project with industrial partner); Excursions to trade fairs and exhibitions, company visits; Competitions (e.g., 24 h challenges, hackathons); Open University/Faculty/Lab days or learning by doing (self-education) in FabLabs and MakerSpaces etc.

On the education market of additive manufacturing, 3D printing there have been observed an interesting dynamic in the last years. After isolated, informal qualification measures from companies for the elimination of a qualification gap (created by new technology) providers of formal advanced training take up topics of the new technology and offer further training in events and courses. Providers may be commercially private, non-profit private or educational institutions of a company, colleges, academies, chambers, guilds or institutions of a church, party, union, or association. Several educational institutions and professional associations have created additive manufacturing certificates, diplomas, and degree programs. Most focus on developing a workforce capable of designing parts, operating the systems, and maintaining the related equipment. There are also various training programs for additive manufacturing, ranging from day seminars to multi-day courses and consulting services.

Aside of companies, there are some relevant examples of public libraries offering 3D printing trainings (together with related services). In 2017, the 3D printing media network has recognized that there were more than 800 3D printers in libraries worldwide and the public needs and interests for this domain of services is increasing [19]; the dynamics of this field of services is positive, related to 2015 when there have been reported 250 3D printers in libraries in the United States supporting the related services to users, according to data published by the American Library Association (ALA). The researchers have found that most printers and services are provided by libraries located in US and other English-speaking countries (e.g., United Kingdom, Australia), but the "Chinese libraries alone could have three to five times as many". Other geographic areas such as South America are also very likely to be offering 3D printing through libraries and European countries, which only recorded a few dozen in this map, are likely to have several hundred (or even thousands) already. Definitively, the Covid-19 pandemic has positively impacted the 3D printing technology use because of the increase interest and the huge number of medical shields, 3D respirator, valves and other devices that were use by hospitals and were donated by public (as universities) or private providers all over the world [20].

2.2 *Industrial Use of 3D Printing*

The exploitation of the 3D printing competencies could be demonstrated by the 3D Printing Index score (calculated based on 38 quantitative metrics in six dimensions) which is an indicator measures the degree to which an array of countries' labor skills, industrial capabilities, governance, and economic assets support additive manufacturing [21]. The index values could shed light not only on where each country stands but also on what can be done to improve the use of 3D printing technologies in actual industries. The Index shows the United States holds the lead with Germany closing in. Korea, Japan, the United Kingdom, and Singapore have also embraced the technology and have begun to capitalize on the opportunity. Challengers such as Australia and China, although lagging, have an opportunity to capitalize on the technology that the leaders have developed, while followers such as Indonesia have a long way to go to catch up (based on data available on the Wohlers Report 2017: 3D Printing and Additive Manufacturing State of the Industry by Wohlers Associates [21]).

Furthermore, the study on the state of 3D printing in 2019 (with 1,300 respondents from Europe (64%), United States (16.6%) and Asia (20.2%), [22]) has underlined that "additive manufacturing adoption is growing across shop floors globally, evidenced by more than 70% of enterprises finding new applications for 3D printing in 2019 and 60% using CAD, simulation, and reverse engineering internally"; 80% of the respondents from enterprises have recognized that 3D printing is enabling them to innovate faster and 51% "actively using 3D printing in production" [22].

From the Eurostat information (data from 2018, published in 2019, see Fig. 1) there have been selected two aspects: (a) the use of 3D printing by economic activity which show a dominant exploitation of the technology and the related competencies by large manufacturing enterprise (more than 200 employees) and (b) the use of 3D printing by purpose that shows the extensive use for prototypes or model for internal production purposes. These two areas of 3D printing competencies exploitation have proven the motivation for teaching and learning in university programs with the support of large-scale manufacturing companies (the situation of successful development) [23].

In the last years, there have been exploited more and more the ideas of creation smart farms or plants of 3D printer networking solution that already exists in the neighborhood areas of universities or big companies' plant. These creative technological hubs will be able to produce more, faster, in better conditions and at lower costs parts for different economic activities (most for the automotive industries, but also for health units, too) [24].

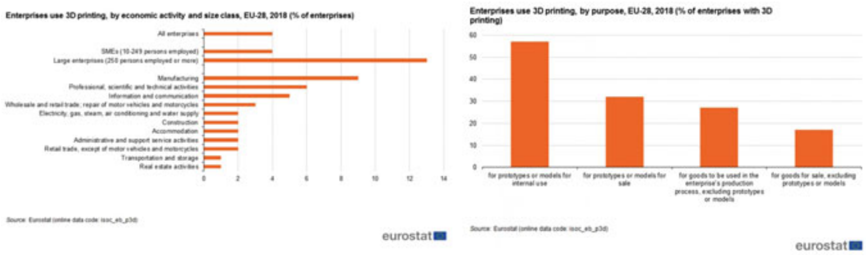


Fig. 1 3D Printing statistics from Eurostat database [23]

3 INNO3D Projects and the Related Training Program

3.1 The INNO3D Project

The present study has been associated with the implementation of the INNO3D project ("3D Printing Support Service for Innovative Citizens") through the support of a European international consortium of libraries, companies, and universities. The project duration is of 36 months, and it started in October 2019, and it aims to develop tools that will define an innovative train program for librarians in using 3D printing. Thus, librarians and specialists (engineers or designers) from their own organization or collaborators (from national organizations or from the INNO3D partners) will be able to train users in using 3D printing effectively and libraries will provide a dedicated service. The project supports the extension of the classical libraries' services [20]. The idea is to transform today's libraries into cutting-edge learning hubs in 3D printing teachnologies and give communities access to technologies that will have a significant impact on fields as scientific research, education, architecture, small scale manufacturing processes for entrepreneurs, engineering studies, healthcare, creative design for culture, entertainment industry (toys) and other [20].

INNO3D international partnership is supported by the following organizations: Limerick Institute of Technology—Ireland (Contractor and coordinator), Transilvania University of Brasov—Romania, Panepistimio Kritis—Greece, Univerzita Konstantina Filozofa v Nitre—Slovakia, Politehnica University of Timisoara—Romania, UNIVERSITAT Politecnica de Valencia—Spain, MB Think Tank srl—Romania, Universidade Nova de Lisboa—Portugal and University of Piraeus Research Center, Greece (nine organizations) [20].

The methodology used for the project implementation is based on six core activities which will deliver the outcomes of the project, as follows: (1) Project management: this is an ongoing activity during the project's life. It ensures the smooth implementation and achieves the proposed results; (2) Mapping 3D printing education in libraries and at the university level; (3) Developing, testing, and implementing 3D printing training curriculum; (4) Elaborating 3D printing trainers' toolkit for theoretical program; (5) Validating quality assurance and improvement of 3D printing curriculum and 3D printing training materials; (6) Transferring project results and best practices. Institutionalizing 3D printing in libraries of each partner in the project to ensure ongoing training of librarians and library users for future generations [20]. The project is expected to have immediate impact and long-term effect on improved and modernized educational processes in 3D printing across university, other libraries in the community and the whole society. Already, the INNO3D training program has been defined based on the inventory of the skills needed in relation to the type of pieces to be process/printed, some multimedia educational resources have been created and currently two e-books are developed one for the trainees and the other for trainers related to the 3D printing technology [20].

3.2 A Proposed Model for 3D Printing Education

Based on the industrial interest on 3D printing competencies and the literature review related to this topic, there have been answered the question on: How 3D printing is being used in the education system? The holistic answer refer to: (1) teach students about 3D printing; (2) teach educators about 3D printing; (3) teach design, creativity skills and methodologies; (4) produce artefacts (pieces, assemblies) to empower the learning process; (5) create assistive technologies (as mention in [2]). Further, the context of INNO3D project implementation has provided new opportunities: innovation transfer related to teaching—learning 3D printing; new skills acquisition in the field of 3D Printing (operating with new and emergent technology); skills related to work and interact in an international project team. Thus, it was created a framework on teaching and education 3D printing, and each partner institutions have benefit from the inspirational ideas of teaching and learning 3D printing of the others. Most, because of the project target group was defined by librarians, the INNO3D project aims to create, through the librarians and the libraries partners (from Ireland, Romania, Portugal, Slovakia and Greece), a critical mass to accelerate the transfer and implementation of 3D printing technologies in engineering education. In the following will be discussed the proposed model adopted by Politehnica University of Timisoara, Romania (Fig. 2).

The case study approach is based on the re-engineering process which is adequate to a disruptive innovation of the library services [25, 26]. It consists of four phases during which several university entities are involved: (1) user-library process to get and understand the order and need of the user; (2) library-CAD model designer (from university specialized department) to prepare the STL AMF 3MF file, do the verifications, repair files if needed, action on orientation load and optimization through simulation; (3) CAD model designer-library process to do the 3D printing process in the library INNO3D MakerSpace; (4) library-user process to deliver the piece and accomplished user need. The framework of "re-engineering model" of the library's services is the result of several discussions with the related stakeholders on the topic.

Fig. 2 The proposed “re-engineering model” of the library’s services Politehnica University of Timisoara, Romania

4 Conclusions

The literature review and industrial trends (before and during the pandemic crisis) have recognized the positive impact and the attractiveness of exploiting additive manufacturing and 3D printing technologies for engineering education. These offers to teaching staff the opportunity to quick demonstrate engineering design process started from ideas (good and bad, innovative one) to artefacts (real models of products) by offering new and challenge opportunities for new teaching—learning—assessment approaches that could be adopt and adapt to a large variety of teaching subjects or learning contexts (e.g., Manufacturing, Production Engineering (innovative manufacturing for product development and Rapid Tooling technologies), Engineering and Technologies (Concurrent engineering, Methodologies and Software tools in Design for Manufacture and Assembly); Biomedical engineering, (Prototypes, customized implants, new biocompatible materials), Architecture, Construction Engineering etc.). Through the case study discussion (the INNO3D project and its implementation at Politehnica University of Timisoara, Romania), there have been underlined the need for university library system in Romania to better support their organization’s change management (adopt the paradigm of library process oriented) strategies to lead them into the present and the future of Digital Era. Considering the proposed feasible approach of re-engineering in university library and their associated digital services (that should be representative and dominant in present and the nearest future), it is necessary to re-think on this serious issue and how it will be applied for providing better services to fulfil multidimensional needs of the users.

The paper main contributions are the fulfilment of the knowledge gap identified in the literature related to the teaching and learning 3D printing in the universities, presenting the advantages, particularities, and the involved actors and by proposing a feasible framework of providing 3D printing technologies as a library service.

Thus, students' projects (mechanical engineering, automotive, logistics, architecture, construction and civil engineering or even architectural models etc.) could easy become a reality and teaching staff could easily access the INNO3D created MakerSpace for demonstrations and practical lessons (e.g., teaching services using INNO3D training program). Also, the MakerSpace is open to industrial applications (external services) or other external users' needs (different applications of students from other universities in the city of Timisoara).

Acknowledgements The paper is linked with the research and dissemination activities related to the "3D Printing Support Service for Innovative Citizens", INNO3D project (2019-1-IE203-000693INNO3D) founded with support of the European Commission. This paper and the communication reflect the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

References

1. Pizzolato, N., Limongelli, M., Di Francesca, A., Kirkar, M., Koch, V., Gomez, V. G., Cruzado, G.C., Fras, M., Rein, E., Silva, R. P.: The "E3D+ VET" Erasmus+ project: interdisciplinary teaching and learning in VET centres through 3D printing. J. Phys. Conf. Ser. **1512**(1), p. 012038. IOP Publishing (2020)
2. Ford, S., Minshall, T.: Invited review article: where and how 3D printing is used in teaching and education. Addit. Manuf. **25**, 131–150 (2019)
3. Pikkarainen, A., Piili, H.: Implementing 3D printing education through technical pedagogy and curriculum development. Int. J. Eng. Pedagogy **10**(6), 95–119 (2020)
4. Zastrow, M.: 3D printing gets bigger, faster and stronger. Nature **578**(7793), 20–23 (2020)
5. Megri, A.C., Hamoush, S.: Teaching high school students innovative topics related to advanced manufacturing and 3-D printing. In: 2020 ASEE Virtual Annual Conference Content Access (2020). https://doi.org/10.18260/1-2-35277
6. Ahmed, W., Zaneldin, E.: The innovation of using 3D printing technology in mechanical and manufacturing engineering. In: IOP Conference Series: Materials Science and Engineering, vol. **938**, no. 1. IOP Publishing, p. 012002 (2020)
7. Colorado, H.A., Mendoza, D.E., Valencia, F.L., Perdomo, J.M.: Production of teaching materials through 3d printing as support for educational processes related to the sciences, heritage, and culture. In: TMS 2020 149th Annual Meeting & Exhibition Supplemental Proceedings (pp. 257–263). Springer, Cham (2020). https://doi.org/10.1007/978-3-030-36296-6_24
8. Wang, X.X., Huang, Z.G., Wen, S.B.: Applications and exploration of CAD-CAE and 3D printing technique in the teaching of plastic products design course. Polym. Bull. **02** (2020)
9. Zaman, M.H.M., Mokhtar, M.H.H., Ibrahim, M.F., Huddin, A.B., Beng, G.K.: Integration of 3D printing in computer-aided design and engineering course. Int. J. Eval. Res. Educ. **9**(4), 934–938 (2020)
10. Rayna, T., Striukova, L.: Fostering skills for the 21st century: the role of Fab labs and makerspaces. Technol. Forecast. Soc. Chang. **164**, 120391 (2021). https://doi.org/10.1016/j.techfore.2020.120391
11. Mersand, S.: The state of makerspace research: a review of the literature. Tech. Trends. 1–13 (2020). https://doi.org/10.1007/s11528-020-00566-5
12. Kinnula, M., Milara, I.S., Norouzi, B., Sharma, S., Iivari, N.: The show must go on! Strategies for making and makerspaces during pandemic. Int. J. Child-Comput. Interact. **29**, 100303 (2021). https://doi.org/10.1016/j.ijcci.2021.100303

13. Szulżyk-Cieplak, J., Duda, A., Sidor, B.: 3D printers–new possibilities in education. Adv. Sci. Technol. Res. J. **8**(24), 96–101 (2014)
14. Lopes, A.J., Perez, M.A., Espalin, D., Wicker, R.B.: Comparison of ranking models to evaluate desktop 3D printers in a growing market. Addit. Manuf. **35**, 101291 (2020). https://doi.org/10.1016/j.addma.2020.101291
15. Romero, M., Lille, B.: Intergenerational techno-creative activities in a library fablab. In: International Conference on Human Aspects of IT for the Aged Population (pp. 526–536). Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58536-9_42
16. Fab lab network. https://fabfoundation.org/global-community/
17. Moeck, P., DeStefano, P., Kaminsky, W., Snyder, T.: 3D printing in the context of Science, Technology, Engineering, and Mathematics education at the college/university level. 21st Century Nanoscience–A Handbook: Public Policy, Education, and Global Trends (Volume Ten) (2020).
18. Sing, S.L., Tey, C.F., Tan, J.H.K., Huang, S., Yeong, W.Y.: Singapore centre for 3D printing (SC3DP), School of Mechanical and Aerospace Engineering, Nanyang Technological University (NTU). Singapore, Singapore (2020)
19. 3D printing media network, there are 800 3D Printers in Libraries Worldwide (and there could be a lot more). https://www.3dprintingmedia.network/800-registered-3d-printers-in-libraries/
20. INNO3D project homepage. https://www.inno3d.eu/
21. Kearnery, Strategic operation articles. https://www.de.kearney.com/operations-performance-transformation/article?/a/3d-printing-disrupting-the-12-trillion-manufacturing-sector
22. Forbes, the State of 3D printing, 2019. https://www.forbes.com/sites/louiscolumbus/2019/05/27/the-state-of-3d-printing-2019/?sh=5b3ac0bd46c2
23. Eurostat database, 3D printing and robotics: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_eb_p3d&lang=en
24. Chepelev, L.L., Rybicki F.J.: The next pandemic and resilience through strategic manufacturing reserves: applying the lessons of COVID-19 in medical 3D printing and other manufacturing. In: Rybicki F.J. (ed.) 3D printing in medicine and its role in the COVID-19 pandemic. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-61993-0_14
25. Koltay, T.: Accepted and emerging roles of academic libraries in supporting research 2.0. J. Acad. Librariansh. **45**(2), 75–80 (2019). https://doi.org/10.1016/j.acalib.2019.01.001
26. Lyon, L.: Librarians in the lab: toward radically re-engineering data curation services at the research coalface. New Rev. Acad. Librariansh. **22**(4), 391–409 (2016)

Multimedia Skills Development to Support Engineering Studies

Anca Draghici, Dana Fatol, Larisa Ivascu, and Muguras Mocofan

Abstract The paper aims describing a collaborative international initiative of four European universities to develop a training program for supporting university teaching staff in creating and using multimedia technologies effectively (extend their digital competencies). Thus, teaching and researching staff from higher education institutions will follow a designed framework to develop their digital skills to create more realistic and attractive content that should improve engineering education (with a debate based on examples and best practices of different specializations). We relate the presented approach to MUST project (Multimedia Competencies for University Staff to Empower University—Community Collaborations, 2020–1-RO01-KA203-080399). The training curriculum and the created educational resources make up a dedicated service offered by universities through DigiCoaches who will provide training to other internal/external users/trainees in creating-using multimedia technologies effectively. The paper will enhance the quality and relevance of knowledge and skills of university teaching and researching staff in multimedia technologies by presenting an overview of the training needs assessment study. Five universities from Romania, Spain, Lithuania, Slovenia, and Republic of North Macedonia with two consulting companies from Germany and Portugal (experts in creating digital resources) will support the new multimedia curriculum and training programme development.

Keywords Education · Engineering · Multimedia skills · Training framework · MUST project

A. Draghici (✉) · D. Fatol · L. Ivascu · M. Mocofan
Politehnica University of Timisoara, Timisoara, Romania
e-mail: anca.draghici@upt.ro

D. Fatol
e-mail: dana.fatol@student.upt.ro

L. Ivascu
e-mail: larisa.ivascu@upt.ro

M. Mocofan
e-mail: muguras.mocofan@upt.ro

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_77

1 Introduction

Universities aim to share and give access to learning and enrichment resources that will have a significant impact on many economic and social fields.

The efforts of universities are beginning to be intense and aligned with the new international requirements. Technological advancement is an efficient support for the development of multimedia technologies. Sound cards led to the development of this technology, followed by several other facilitators: compact discs, digital cameras that began to be used more and more, and finally the various videos that turned the computer into an efficient educational tool. Multimedia technologies are used in various environments: medical, statistical, economic and others. At the same time, multimedia technologies are used in entertainment and in the active training of individuals. In recent years, multimedia technologies have been very good facilitators of the teaching–learning process. It is appreciated in the literature that these technologies are the best educational techniques. This statement is supported by the following aspects: it addresses several senses and develops the creative side. All these presentations used in universities integrate several elements that develop an interest in students: text, spoken words, sound, music, graphics, animations, photographs, images, numerical evolutions, and others.

New multimedia technologies will transform today's universities into cutting-edge learning hubs that teach 21st-century skills. The goal of multimedia technologies in universities is to create a realistic perception of the users on novel learning technologies and potential of digital resources and reinforce that learning with simpler hands-on projects [1–8]. The advantages of using multimedia technologies are many that can be remembered: transforming the reading process into a dynamic process, the use of graphics helps to shape the concept and faster learning, the recorded voice can be heard several times so that the concept is enhanced, various colours and animations can turn the learning process into an innovative and attractive one, and the use of maps and classifications outlines a concept closer to reality [9–14].

The study [14, 15] presents the impact of multimedia technologies on universities in Asian regions and highlights the main factors identified: efficiency and attractiveness. Other studies are important in the field of mathematics in which remarkable results are recorded [16]. In chemistry and biology, universities that use multimedia technologies have favourable results, especially in engineering [17–20]. There were also studies in universities that have study programs in management. Here there was a much higher degree of attractiveness and a considerable presence at courses. Engineering universities have started to use these facilities more and more intensely, and the results show an increased interest of students [21]. Engineering and the teaching process are directions in which new technologies are present and represent a basis for future research directions [22, 23]. Therefore, multimedia technology is a growing interest in universities, but also for the economic environment [24–26].

This paper aims to present an international initiative for the development of a blended curriculum and training program dedicated to university staff to support their multimedia skills development. To develop this approach, a market research survey was conducted to identify multimedia training needs. The proposed approach will support engineering studies from several European universities.

2 Overview of MUST Project

MUST project (Multimedia Competencies for University Staff to Empower University—Community Collaborations, 2020–1-RO01-KA203-080399) addresses to "Innovative practices in a digital era" (as described in [20], p. 101), because there is proposed to be develop-implement an innovative blended curriculum and training program dedicated to university staff to support their multimedia skills development; these refer to the design and use of digital technologies in creative, collaborative, efficient ways to better facilitate teaching-training-learning-assessment process of Generation Z students or for user in universities' community (vocational education for lifelong learning). These are aligned with the objectives of the European Education Area, notably the new European Universities initiative [21]. Considering the European Commission Digital Education Action Plan [22] that calls for systematic action in this area, MUST project addresses to the developing of digital competences and skills (Actions 4 to 8) and the use of the European frameworks on digital competences of educators as university staff.

In each university partner involved in the project there were not implemented coherent training programs for multimedia skills development of teaching staff (as will be demonstrated through the training needs assessment). The developed high quality visual digital media resources and learning framework can help to engage this target group in meaningful learning activities for self- expression and gradual competence and confidence building in key competences as well as the technical skills to use sector specific occupational learning. Visual media within a transmedia strategy make learning independent of time, allow for repeating the content as needed and can be designed to promote language proficiency, digital skills, a sense of initiative and self-efficacy at the same time. Common projects using engaging multimedia with university staff and students/community learners can lower the mental barriers for accessing higher education [19].

The project is developed transnationally, because none of the participating university has enough expertise available to carry out the programme on its own (as demonstrated also by the Pilot Study, see Sect. 3 and 4 of the paper) and they need support from expertise of an NGO partner from Germany (SoWiBeFo) and a small company from Portugal (Storytellme). Benchmarking expertise on work based learning and digital media developed in Germany (referring to universities and industry) and experiences from the use of text-based digital learning in Portugal will enriched and extend university staff knowledge in the design and use of multimedia technologies [19].

The project objectives are to improve and diversify: (1) university staff skills in creating and using multimedia technologies effectively (for education, research, university-community projects, advertising, and communication purposes); (2) the services offered to university users and other university staff through the designed training program that empower these users with key skills and the opportunity to develop their multimedia competencies. Following MUST training program, skilled university staff could better support not only internal processes of the universities but also university-community project by creating a strong imagine (through visual storytelling, experiments, and demonstrations, through transmedia etc.) on the academic potential for knowledge-innovation transfer, lifelong learning and vocational education training, social and cultural projects, university advertising etc. The developed MUST training curriculum and the created educational resources will define a dedicated service offered by universities through DigiCoaches (trainers of trainers from university staff) who will provide training to other internal/external users/trainees in creating-using multimedia technologies effectively.

3 Research on Multimedia Training Needs Assessment

A survey based on a designed questionnaire was conducted supported by the Google Form platform; the answers were collected from January 2021 to June 2021. This study was applied online and addressed to individuals from the universities from Romania, Germany, Lithuania, North Macedonia, and Slovenia. The research sample consists of all types of universities staff (teaching, administrative, technical, students); 122 responses were registered, being representative for the targeted university. From the perspective of age, there are respondents in the category under 25 years, approximately 27%, the interval between 25 and 35 years there were 15%, the interval 35–45 years a percentage of 30%, and over 45 years a percentage of 26%. From a gender perspective, there is a well-balanced proportion, about 53% women, and 47% male. The areas in which the respondents work, and they will exploit the achieved multimedia knowledge is manly from the engineering field of science (51.70%).

Figure 1 shows the results regarding the respondents' interest on several topics to get involved in such training programs. The skills held are suitable for the job, but they can be improved. Respondents are willing to improve their multimedia skills because they can strengthen both research and practical skills.

Figure 2 highlights the respondents' training needs in multimedia production and exploitation fields. There is interest in a "strong" and "very strong" need for knowledge exploitation. From the general basic knowledge perspective, the directions targeted by this research were: relevant intellectual property rights in the context of multimedia design, production and exploitation, planning strategies for multimedia design, multimedia production techniques, and methods and tools used in multimedia production. In each direction, multimedia applications in university educational processes are appreciated by respondents and can be used in educational programs, research, communication, and others. In addition, the preferred method for achieving multimedia skills and competence has been investigated. Over 46% appreciate that best practices represent the main direction. Tutorial demonstrations are appreciated by 31% of respondents.

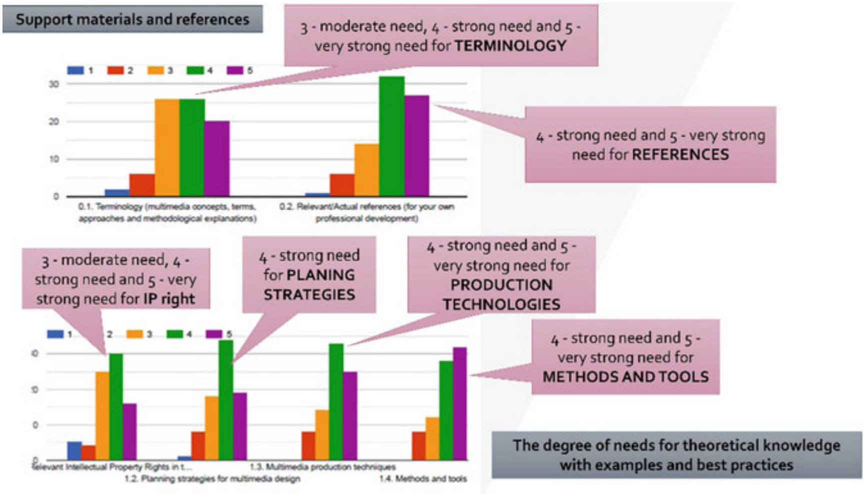


Fig. 1 The responses reflecting the need for "Support material and Reference" and for theoretical knowledge in different areas

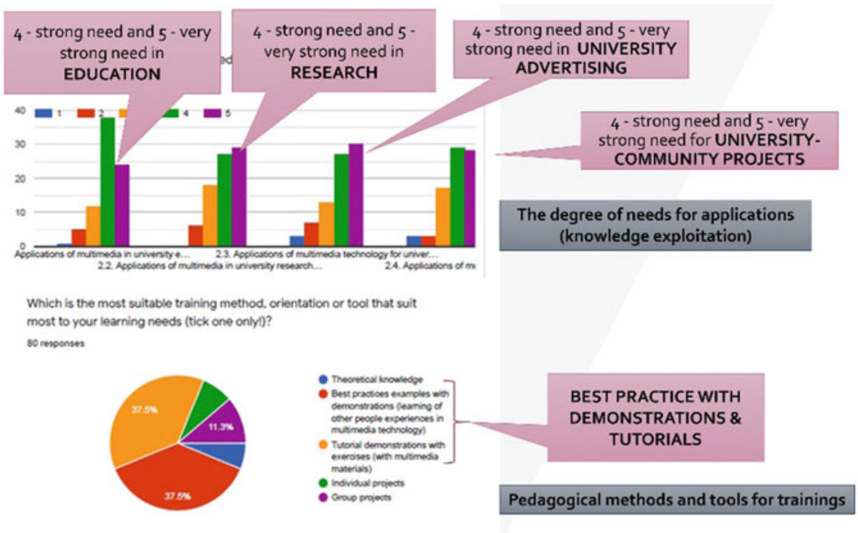


Fig. 2 Respondents' interest in getting involved in a program for developing multimedia skills

4 Multimedia Skills Development to Support Engineering Studies

The training needs assessment has provided a large range of information regarding the needs of multimedia skills and competencies. During the pandemic period (started from March 2020 till June 2021) several informal discussions with colleagues involved in the education process of different engineering programs (specializations) have underlined that many of them have self-developed their digital skills in teaching on-line. Difficulties encountered when teaching practical lessons (demonstrations on handling and using equipment, apparatus, machines etc.), most of them being forced to quick develop movies in the university labs, providing realistic visual demonstrations that were used in the online classes. Most of the teaching staff were not equipped with properly recording equipment and thus most were using their mobile phones for these practical lessons. Frequently, teachers have no knowledge to do the recording and the video explanations (they were not able to provide clearly and attractive movies due to their lack of acting skills) and the results usually were supported by additional explanations with other video demonstrations found on YouTube or provided by industrial consulting companies. In addition, during the online classes via Zoom, the generated movies need to be process and re-use adequately for students learning process. So, additional problems arise in the field of self-processing of the video theoretical lessons and the tutorials (related to the engineering seminars and projects development). Simultaneously, usually in the universities departments there are no technical staff able to support multimedia materials (e.g., educational resources).

Started from this situation and the need for quick extend the teaching staff digital knowledge in the field of multimedia technologies arise in the early period of the pandemic when the MUST framework for multimedia competencies development together with the associated training program have been designed in a collaborative manner within the international consortium.

The proposed framework for multimedia skills development has been developed based on the Bloom's taxonomy (adapted for https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/). The adaptation takes into considerations five levels of the taxonomy by merging level 1 and 2 of the original shape due to the way learning and action (knowledge use and exploitation) in the case of multimedia technologies required such an approach. According to this framework, a preliminary self-assessment test provides a diagnosis, feed-back of the potential trainers' skills and competences (done online via the web page and the e-learning platform dedicated to MUST training program) to the design and development of the multimedia products (see Table 1).

In addition, the self-assessment test offers concrete suggestions for competence improvement in a specific domain! Moreover, university staff will get the opportunity to increase their competence level by enrolling in the corresponding online training program that will be developed by MUST consortium. Depending on the identified proficiency level, potential trainers might be suggested to complete either the entire training or only specific unit(s). Each area is assessed independently; individuals

Table 1 The structure of the self-assessment test defining MUST framework for multimedia skills development

Area of competences to be assess (no of competences/training elements)	Evaluation scale (applicable for all the training elements developed in correlation with the developed competences)	Test results and recommendations
(U1) Intellectual Property Rights for multimedia (2 competences/training elements)	• I have basic knowledge and hear about …; I can explain the basic principles of … • I can design a multimedia product and apply knowledge for … • I developed analysis and synthesis based on the knowledge in the field of …; I can design a approach for … • I can evaluate … • I can create professional … • No statement is correct	• Feed-back for each training element • General feedback per training unit • General suggestion on which element and unit should be followed
(U2) Planning strategies for multimedia design (2 competences/training elements)		
(U3) Multimedia production techniques (2 competences/training elements)		
(U4) Methods and tools for multimedia production (4 competences/training elements)		
(U5) Applicative competences (education, research, advertising, community projects)—four domains each with 2 competences/training elements		

although can freely choose any area, he/she should assess the five areas sequentially, using the same scale as suggested by the Blooms' taxonomy (individuals should choose the option that best fits to his/her competence level). After the self-assessment is finalized, potential trainers will be given general feedback in terms of his/her competence level for the whole area (unit) and each related training-learning element (which correspond to a specific competence that is developed) in certain cases he/she will be pointed to the right training modules that should be followed to further develop his/her competence level.

The described framework is correlated with the training program skill card (first column of Table 1). The total number of competencies is 18 and they refer both to the Pedagogy and Technology domains (as identified by MUST consortium members); these were considered most relevant for the creation and delivery of the MUST training program for university staff (and will allow the transformation into a post-university training program). The defined competencies (associated to the training elements of each unit) were also verified by participants in 2 focus groups consisting of academic staff and instructional designers from partner institutions in Slovenia, Lithuania, Portugal, Germany, Spain, North Macedonia, and Romania. Thus, MUST training program is already structured for the university staff (including teaching, research, and administrative staff) multimedia skills development. In the next period there will be developed the in Europe and beyond will receive accurate input on the development and implementation of MOOC-based curricula. MUST training program itself will be designed as a MOOC in English language and it will be translated in all university partners languages. In this way, the wider international academic community will be granted with open access to this output and will learn about multimedia design and technologies.

The contents of the training program will be closely linked to the self-assessment tool. Based on the results obtained from the self-assessment tool, academic staff will be re-commended to either take separate modules of the training program, or to complete the entire program.

5 Conclusions and Final Remarks

The pandemic situation has changed the way we approach the learning process. In this context we must speak of learning, unlearning, and relearning. Competency building needs to be strengthened through multimedia technology training programs. The MUST project is a necessity for technical universities in different countries. The training needs were identified through an online market research conducted in individuals in Romania, Germany, Lithuania, North Macedonia, and Slovenia. Respondents appreciate the practical studies and simulations that can be performed using these technologies.

Strengthening multimedia skills is a necessity in these pandemic conditions where online learning has replaced the traditional face-to-face learning. University teaching staff skills' development in the field of multimedia can attract students in the teaching process and increase the motivation rate (develop active learning communities). Also, it is expected that the proposed approach will positively impact students' way of communication and better expressed themselves.

The presented framework will be the concerned of the next period activities of the MUST project. Thu, a study on the generated effects of the learning program is needed to really proof the impact of the achieved skills in engineering studies (promoting best practices, successful pedagogical methods etc.).

Acknowledgements The paper is linked with the research and dissemination activities related to the MUST project: "Multimedia Competencies for University Staff to Empower University—Community Collaborations" (Erasmus+ 2020-1-RO01-KA203-080399), founded with support of the European Commission. This paper and the communication reflect the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

References

1. Al-Ajmi, N.H., Aljazzaf, Z.M.: Factors influencing the use of multimedia technologies in teaching English language in Kuwait. Int. J. Emerg. Technol. Learn. (2020). https://doi.org/10.3991/IJET.V15I05.12277
2. Akor, T.S., Bin Subari, K., Binti Jambari, H., Bin Noordin, M.K., Onyilo, I.R.: Engineering and related programs' teaching methods in Nigeria. Int. J. Recent Technol. Eng. (2019). https://doi.org/10.35940/ijrte.B1915.078219
3. Balcita, R.E., Palaoag, T.D.: Augmented reality model framework for maritime education to alleviate the factors affecting learning experience. Int. J. Inf. Educ. Technol. (2020). https://doi.org/10.18178/ijiet.2020.10.8.1431
4. Ding, T., Zhu, K.: The application of multimedia technology in social media content marketing. J. Phys: Conf. Ser. (2020). https://doi.org/10.1088/1742-6596/1648/3/032052
5. Dlaska, A.: Sites of construction: language learning, multimedia, and the international engineer. Comput. Educ. (2002). https://doi.org/10.1016/S0360-1315(02)00031-3
6. Gluzman, N.A., Sibgatullina, T.V., Galushkin, A.A., Sharonov, I.A.: Forming the basics of future mathematics teachers' professionalism by means of multimedia technologies. Eurasia J. Math. Sci. Technol. Educ. (2018). https://doi.org/10.29333/ejmste/85034
7. Khazaal, H.F.: Problem solving method based on E-learning system for engineering education. J. Coll. Teach. Learn. (TLC). (2015). https://doi.org/10.19030/tlc.v12i1.9066
8. Klemeš, J.J., Kravanja, Z., Varbanov, P.S., Lam, H.L.: Advanced multimedia engineering education in energy, process integration and optimisation. Appl. Energy (2013). https://doi.org/10.1016/j.apenergy.2012.01.039
9. Jahansouz Shahi, M.A.: The impact of E-learning on improving Iranian EFL learners' language skills: decreasing learning anxiety. J. Fundam. Appl. Sci. (2016). https://doi.org/10.4314/jfas.v8i3s.180
10. Jajuri, T., Hashim, S., Ali, M.N., Abdullah, S.M.S.: The implementation of science, technology, engineering and mathematics (Stem) activities and its effect on student's academic resilience. Asia Pac. J. Educ. Educ. (2019). https://doi.org/10.21315/apjee2019.34.8
11. Juan, L., Yahaya, N.B.: The problems and countermeasures of applying multimedia technology in college English teaching. Int. J. Eng. Adv. Technol. (2019). https://doi.org/10.35940/ijeat.E1222.0585C19
12. Indra Devi, S., Feroz, F.S., Kasturi, K., Saad, S.M., Khoo, M.S.L.: Multimodality in education 4.0: a case in an engineering context. Int. J. Adv. Trends Comput. Sci. Eng. (2020). https://doi.org/10.30534/ijatcse/2020/267942020
13. Inovasi Dalam kejuruteraan mekatronik: perbandingan pencapaian pelajar yang berorentasikan kemahiran analisis Dengan Berorentasikan Sosial. J. Techno Soc. (2009)
14. Li, Y., Lu, S.J.: Research on physical education system model using multimedia technology. Multimedia Tools Appl. (2020). https://doi.org/10.1007/s11042-019-08366-7
15. Liu, Q., Chen, H., C. Crabbe, M.J.: Interactive study of multimedia and virtual technology in art education. Int. J. Emerg. Technol. Learn. https://doi.org/10.3991/IJET.V16I01.18227
16. Liu, X., Liu, Y., Tu, J.F.: Multimedia technology and learner autonomy: an experimental study for asymmetric effects. Symmetry (2020). https://doi.org/10.3390/sym12030462
17. Park, C.W., Kim, D.G., Cho, S., Han, H.J.: Adoption of multimedia technology for learning and gender difference. Comput. Human Behav. (2019). https://doi.org/10.1016/j.chb.2018.11.029
18. Prayitno, T.A., Hidayati, N.: Multimedia development based on science technology engineering and mathematics in microbiology learning. JPBIO (J. Pendidikan Biologi). (2020). https://doi.org/10.31932/jpbio.v5i2.879
19. Rakhimova, A.E., Mukhamadiarova, A.F.: Multimedia technologies as an educational environment in the formation of sociocultural competence. Opcion (2020)
20. Ren, X., Cui, J.: The development and application of multimedia technology in college gymnastics teaching. J. Test. Eval. (2021). https://doi.org/10.1520/JTE20200196
21. Shi, X.: Application of multimedia technology in vocabulary learning for engineering students. Int. J. Emerg. Technol. Learn. (2017). https://doi.org/10.3991/ijet.v12i01.6153

22. Shu, Y.: Experimental data analysis of college English teaching based on computer multimedia technology. Comput. Aided Des. Appl. https://doi.org/10.14733/cadaps.2020.S2.46-56
23. Tsai, C.H., Kuo, Y.H., Chu, K.C., Yen, J.C.: Development and evaluation of game-based learning system using the microsoft Kinect sensor. Int. J. Distrib. Sens. Netw. (2015). https://doi.org/10.1155/2015/498560
24. Yao, Y., Wang, P., Xia, X., Li, X., Song, C.: The application of multimedia technology in teaching innovation. J. Test. Eval. (2021). https://doi.org/10.1520/JTE20200232
25. Vaganova, O.I., Bakharev, N.P., Kulagina, J.A., Lapshova, A.V., Kirillova, I.K.: Multimedia technologies in vocational education. Revista Amazonia Investiga (2020). https://doi.org/10.34069/ai/2020.26.02.45
26. Zarei, A., Yusof, K.M., Daud, M.F.B., Kolivand, H., Saba, T.: A novel approach of multimedia instruction applications in engineering education. J. Theor. Appl. Inf. Technol. (2016)

Healthcare Management

An Optimization Model on Patient Appointment Scheduling of MRI Diagnostic Examination with Prioritization

Joachim Jover, Westin Perez, Joanne Tung, and Charlle Sy

Abstract Purchasing additional MRI machines may not always be a feasible solution to increase diagnostic services capacity. Healthcare systems make use of appointment scheduling to optimize the use of their limited resources as an alternative to investing in new machines. A properly planned appointment scheduling system will allow for the efficient use of the MRI machine and accommodate patients effectively. The challenge in appointment scheduling comes in the various uncertainties that could disrupt the planned schedules. Current research and literature on appointment scheduling for MRI scans have not tackled the appointment scheduling problem while considering the occurrence of unpunctual patients, no show patients, and unscheduled arrivals, at the same time. This study develops a scheduling model that considers the aforementioned uncertainties in addition to different patient types. A two stage solution is utilized in order to create a weekly appointment schedule. The first stage determines the total number of patients to be accommodated for the week. This is then used as an input for the second model to determine the time slot assignment of patients. The latter minimizes the total penalty contributed by the priority levels and the cost of rearranging from the original order of requests. A neighborhood search heuristic is used to develop a prioritization rule when such uncertainties are realized. From the results and analysis, it was found that the model returns a schedule that generates a lower value of the objective function compared to ad hoc scheduling practices such as that of scheduling according order of request only or according to priority level only.

Keywords Optimization · Patient scheduling · Healthcare · Heuristic

1 Introduction

Capacity deficiencies occur when the demand for a service cannot be met using available resources. The Diagnostic Demand and Capacity Programme [1] reports

J. Jover · W. Perez · J. Tung · C. Sy (✉)
Industrial and Systems Engineering Department, De La Salle University, Manila, Philippines
e-mail: charlle.sy@dlsu.edu.ph

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_78

that MRI capacity is unable to match demand and yet the utilization rate of the equipment only averages from 43.7 to 59.8%. This may be caused by both the lack of equipment and the varying protocols of the hospitals. Low utilization rates of the MRI machines are still observed despite demand being higher than capacity. This highlights the significance of proper planning of resources in order to satisfy the patients' needs without the need for additional investment. A study by [2] states that MRI examination is one of the greatest bottlenecks in hospitals since capacity is limited.

Hospitals need to operate more efficiently and effectively with the limited resources they have. This was addressed in [3] by highlighting the importance of hospital scheduling since proper scheduling of operations improves staff productivity and morale. This in turn improves the efficiency in the utilization of resources and equipment. Appointment scheduling, if applied properly, allows for the system to undergo a smooth workflow, reduces waiting time of patients while lessening overcrowding, and provides the ability to match supply and demand in the current system.

However, proper application of appointment scheduling is no easy task, as difficulty in scheduling is often caused by the inconsistency of the demands hospitals face, given the variety in the patients' diagnosis and the equally varied urgency these patients need for treatment [4]. Studies have been performed on managing hospital appointment schedules, tackling a variety of the problems that arise from the procedure. Mahmoudzadeh et al. [5] tackled the uncertainty in the demand for appointments, given its unpredictable nature in a hospital setting. In a separate study, [6] looked into optimizing the appointment scheduling process by accounting for same-day requests and "no-show" patients that fail to show up for their scheduled appointments, as this is a reality that hospitals face every day and can cause problems to arise if not accounted for [7]. Tackled similar ideas in their study, but opted to also account for an outpatient hospital setting in their study, as well as delving into unpunctual arrivals and stochastic service times when searching for ways to optimize scheduling. However, the study does not consider unscheduled arrivals or same-day requests.

Despite the substantial amount of research that covers multiple types of patients regarding their appointment schedules and whether or not they arrive for them, no study has yet to account for all types of patients. These types being: on time patients, late arrivals, no shows, and unscheduled patients in a single system. Patient unpunctuality is a prevailing issue when tackling MRI scheduling, as lates and no-shows are very common among MRI patients. A study performed by [8] determined that in particular, unpunctual patients serve as a major detriment to the MRI scheduling process, with no-shows proving to be particularly detrimental, given that these patients lead to slots being wasted on them and double occupancy for MRI scheduling slots.

The results that previous models and studies arrive at are not applicable to a model that seeks to incorporate all types of patients due to the fact that no-shows and late patients ultimately provide different types of problems that the system must deal with. This is due to the fact that late patients serve as delays in the system, as patients who are late do still make use of the system beyond their initial intended arrival time, while no-show patients essentially free up slots in the system, providing opportunities to rearrange patient schedules in more optimal positions. This is something that this study seeks to cover, in order to create a more comprehensive analysis of the hospital system by accounting for a variety of patient accommodation concerns. Since there is a high demand for MRI, hospitals use scheduling as a means to organize the machine's activity throughout the day. Through appointment scheduling, the hospital will be able to ensure that more patients will be given treatment on time and that the machine is better utilized with less idle time. An appointment scheduling model that considers various patient accommodation concerns will be able to generate a schedule that can adjust when uncertainties are realized.

2 Appointment Scheduling for MRI Diagnostic Exams

The components of the scheduling system for MRI examinations consist of the following: the patients, the scheduler of the hospital or MRI clinic, the MRI machine, and the initial schedule of patients. The schedule appointed to patients will depend on the patient's priority level and the order within the queue. Priority level pertains to the urgency or level of care needed by the patient as assessed by the doctor's recommendation to have an MRI diagnostic test. The order of the patients is the original sequence of the patients' request for an appointment. This is likened to a first come first served queue wherein those that request earlier are placed first in the order of appointments. The priority level of the patients would affect the order of the patients if higher priority patients are present in the system and they have to be accommodated before those that made an appointment request earlier. An appointment time will be given to the patients however, the given schedule for patients within a day may still be subjected to uncertainties.

The number of patients that can be accommodated into the system is based on the capacity of the available MRI machine. For appointment booking, the scheduler will contact the patient first before they are assigned to a schedule. During this first call, the scheduler will obtain the patient's information before placing them in the queue of exams to be scheduled. If during the first call attempt, the necessary patient information has not yet been obtained, the patient drops the call, or the patient cannot be contacted, a second call attempt will be made within one week of the first call. Based on the aforementioned study, the minimum time patients spent in the queue was 2.5 weeks. This process shows the current set-up for scheduling an appointment. Comparing the access time of the patient from the current setup of 2.5 weeks lead time, a mathematical model will significantly reduce the access time because patients can be scheduled instantaneously. This will help determine the approximate number of patients that the facility has to prepare for in a given week. The appointment booking flowchart used in the study is shown in Fig. 1.

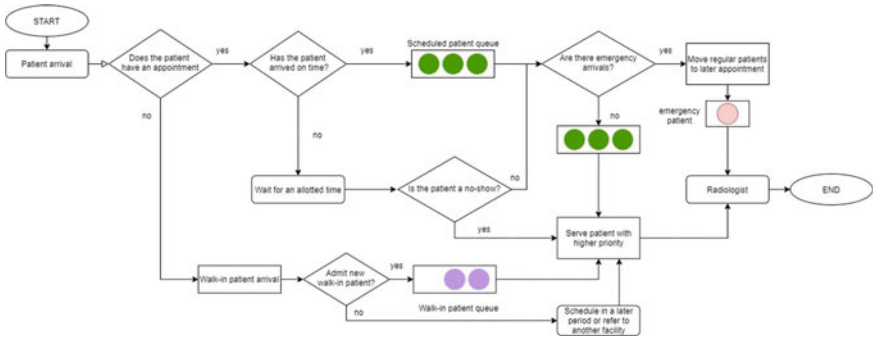


Fig. 1 Intraday schedule flowchart

3 Model Development

The patient appointment scheduling system for MRI exams identifies the number of patients to expect in a given week and the patient schedule to assign and prepare throughout the workweek. As a result, the study opted to use two models to determine these separate goals. The first model will create an estimate of the number of patients that must be prepared for in a workweek. This would allow for the hospital to aptly prepare for the expected patient distribution and understand the size of the schedule in that week. This is where the second model would come in, where with the knowledge of the number of patients expected in the workweek, there is an understanding of how the model should be utilized.

Prior to running the second model, the number of runs required is first computed. This is done by dividing the number of patients to be accommodated for a week, obtained from the first model, by the number of patients that can be scheduled in one run of the second model. This means that the number of runs for Model 2 is affected by the size of the model that can be run within a reasonable amount of time. After computing for the number of runs, the initial order of patients and their corresponding priority levels can be inputted to Model 2. The second model will create a schedule that minimizes the schedule rating, based on the initial schedule and the priority levels. Once a schedule has been formulated, a heuristic will be used to determine a prioritization rule when no shows or unscheduled arrivals occur. The neighborhood search heuristic will be used to make the adjustments to the model when information on these uncertainties are realized. This will result in a finalized and adjusted schedule considering uncertainties.

3.1 Total Patient Prediction Model

The first model formulated for the study is a model designed to predict the amount of patients that the hospital system must account for based on probability of lates and no shows, the expected amount of emergency patients, and the expected distribution of patient priority types. This model will be run first given that the output will serve as an input for the second model. For simplification purposes, we will refer to the first model as Model 1 throughout the study.

Equation (1) represents the objective of the model, which is to prepare for a realistic number of patients of varying priority levels. The model was created to set up a scenario where hospitals will maximize the number of patients they can accommodate. Thus, Z_a is defined as a summation of the weight distributions garnered as a result of taking in patients of various types (x_{ij}), with an altered weight contribution for accommodating beyond hospital capacity. The decrease of weight contribution as the model allocates patients above the capacity limit is given such that an optimal schedule is one that follows the idea that the more patients there are, the higher the waiting time for the patients above the limit. This can also be interpreted as congestion in the hospital system or overcapacity. The penalties then represent the costs that are incurred by accommodating patients beyond the soft limits of the hospital setting.

$$Max Z_a = \left(\sum_{i=1}^{n} Value_i * \sum_{j=1}^{n} x_{ij} \right) - \left(\sum_{i=1}^{n} Penalty_i * x_{i2} \right) \tag{1}$$

Equation (2) establishes the distinctions between patients that fall within the hospital limits when they are accommodated and those who do not. This constraint is necessary in order to establish the basis for cost increasing per patient accommodated beyond this limit. In order to make the distinction between patients that fall within the soft limit and those outside, a switch variable Y is used as in Eq. (3) to create a condition where patients outside the soft limit can only be accommodated once the initial soft limit was filled up.

$$\sum_{i=1}^{n} x_{ij} \leq Limit_j \, \forall j \tag{2}$$

$$\sum_{i=1}^{n} x_{i2} \leq M * Y \tag{3}$$

To create a projected allocation amount for patients of a specific type, the total amount of patients of said type was constrained to fit within certain percentage parameters. This would create a scenario where the model would expect to fit each amount of patients in a specific priority type into an allocated portion of the whole, resulting in a projected value represented by $Allocation_i$. Once the allocation amount is determined, this value will then be subject to slight modifications based on punctuality probability as well as emergency patient probabilities. This would mean that the actual amount of patients the model prepares for will not be equal to the allocation amount previously determined, as the resulting value must still be based on no-show, late, and emergency probabilities.

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} * Distribution_i \leq Allocation_i \quad \forall i \tag{4}$$

$$\sum_{j=1}^{n} x_{ij} + Emergency_i = Allocation_i * Punctual_i \quad \forall i \tag{5}$$

3.2 Patient Scheduling Model

Based on the results of Model 1, the Patient Scheduling Model (referred to as Model 2) will then create an optimal schedule based on initial appointments and patient priority levels. The model's objective function, shown in Eq. (6), minimizes the combination of the total weighted contribution of patients in the system based on their position in the schedule and the cost accrued by adjusting patients from their initial positions in the process of creating said schedule. Patient weight contributions are computed by multiplying their predetermined priority levels with their position on the final schedule, where the patient that goes first would be assigned a value of one (1), the patient that goes second would be assigned a value of two (2), and so on. This would provide incentive to place higher priority patients first in the schedule as to minimize the multiplicative values weight contributions provide. Without the penalties in the model, the optimal solution that would be generated arranges patients by priority level in descending order. In a practical setting, these penalties translate to a decrease in the patients' satisfaction with the service. If those with lower priorities are appointed at the end of the schedule despite having requested earlier, then their satisfaction with the service decreases.

$$MinZ_b = \sum_{k=1}^{n} (Priority_k * x_k + RatingDecrease_k * Change_k) \tag{6}$$

Equations (7) and (8) create unique schedules for each patient in the system. Hence, a switch variable was used to determine the greater value between x_k and x_l when compared to one another. It is also important to ensure that each patient has a minimum space of one unit of schedule between them. Meanwhile, Eq. (9) ensures that all patients have an appointment in the hospital system, preventing patients from not being listed in the resulting schedule. In Eq. (10), the absolute value of the difference between the patients' original and new positions will be multiplied to the corresponding penalty weight to account for adjusting patients from their original appointment schedules.

$$x_k - x_l \geq 1 - M * Y_{kl} \quad \forall k, l \tag{7}$$

$$x_l - x_k \geq 1 - M * (1 - Y_{kl}) \quad \forall k, l \tag{8}$$

$$x_k \geq 1 \quad \forall k \tag{9}$$

$$|x_k - Original_k| = Change_k \quad \forall k \tag{10}$$

3.3 Heuristic

A neighborhood search heuristic is subsequently used to account for uncertainties that might happen once an initial schedule has been obtained. In a real life setting, when uncertainties such as no shows or unscheduled arrivals occur, the scheduler will have to make the necessary adjustments as soon as possible.

The mechanism for Algorithm 1 and Algorithm 2 are shown in Fig. 2. Algorithm 1 defines the process of updating the schedule when there is a no show patient. The process begins when a no show patient is realized. The scheduler will then have an option to call on a patient from the set of remaining patients who have yet to be serviced for the day. The criteria for choosing which patient to be moved up to the time slot of the no show is either to call the immediate next patient or the next patient with the highest priority level among the remaining patients. If the patient can be moved up to the earlier time slot, the scheduler will then assign them to the no show patient's timeslot starting after the 15 min allocated to the lateness allowance. However, if the patient is unable to take the earlier time slot, the schedule of the remaining patients will continue as is.

Algorithm 2 defines the process of updating the schedule when there are unscheduled arrivals. The process begins when an unscheduled patient arrives. The priority level of the unscheduled arrival is checked and compared with the priority level of the set of remaining patients for the day. The criteria to determine which time slot the unscheduled arrival will take is either to schedule the new patient at a time slot where all scheduled patients of a higher priority have been served or to schedule the new patient at a time slot where all scheduled patients of the same priority level have been served. After the new patient is given a time slot, the scheduled patients will then be pushed back by the number of unscheduled patients allocated before them. If there are no unscheduled arrivals, the schedule will proceed as is.

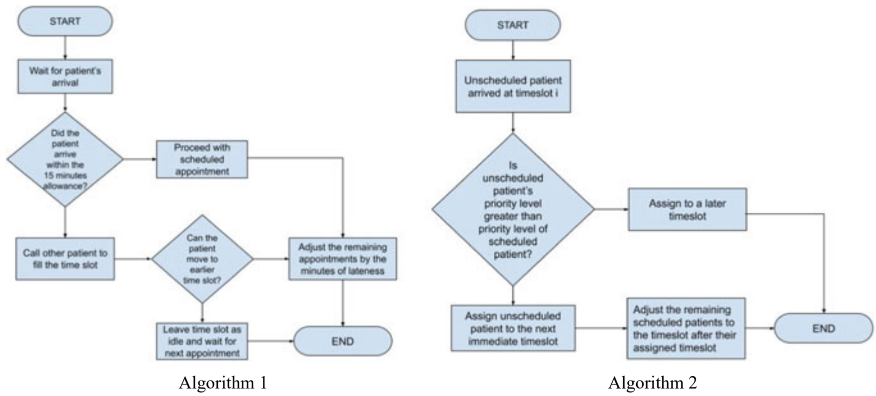


Fig. 2 Flowchart

4 Model Results

The model optimization process involves the calculation of the amount of patients the hospital system should prepare for and the identification of the order of the aforementioned patients with consideration of their priority level while limiting the amount of change in individual patient schedules from their original requested timeslot. Figure 3 shows the results gathered from Model 1 and Model 2. A total of 157 patients, belonging to four priority levels (1 being the lowest and 4 having the highest priority), were assumed to be needing an MRI schedule. The higher priority patients were naturally the ones that fit within the soft limit. Lower priority patients were accommodated beyond the limit as much as the system could allow it. Soft limit is the capacity of the hospital which can be flexible.

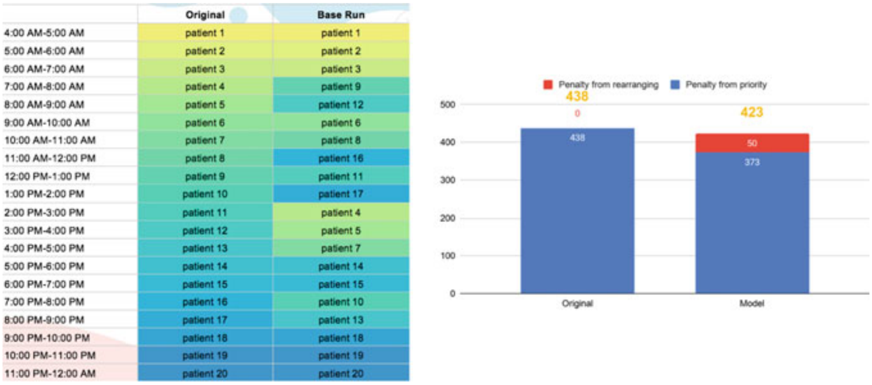
Model 2 gives the output of column 'X' and 'Timeslots moved'. 'X' denotes $\mathbf{X_i}$ which is the order of patient numbers. For example, Patient Number 4 should be placed in the 11th time slot because it has a corresponding **Xi** value of 11. The

Model 1

X	Patient Number	Priority Level	New Order	New Priority	Timeslots Moved
1	1	4	1	4	0
2	2	2	2	2	0
3	3	4	3	4	0
11	4	1	9	4	5
12	5	1	12	3	7
6	6	2	6	2	0
13	7	1	8	3	1
7	8	3	16	3	8
4	9	4	11	3	2
16	10	1	17	3	7
9	11	3	4	1	7
5	12	3	5	1	7
17	13	1	7	1	6
14	14	2	14	2	0
15	15	1	15	1	0
8	16	3	10	1	6
10	17	3	13	1	4
18	18	1	18	1	0
19	19	2	19	2	0
20	20	2	20	2	0

Model 2

Fig. 3 Results from model 1 and model 2



	Original	Base Run
4:00 AM-5:00 AM	patient 1	patient 1
5:00 AM-6:00 AM	patient 2	patient 2
6:00 AM-7:00 AM	patient 3	patient 3
7:00 AM-8:00 AM	patient 4	patient 9
8:00 AM-9:00 AM	patient 5	patient 12
9:00 AM-10:00 AM	patient 6	patient 6
10:00 AM-11:00 AM	patient 7	patient 8
11:00 AM-12:00 PM	patient 8	patient 16
12:00 PM-1:00 PM	patient 9	patient 11
1:00 PM-2:00 PM	patient 10	patient 17
2:00 PM-3:00 PM	patient 11	patient 4
3:00 PM-4:00 PM	patient 12	patient 5
4:00 PM-5:00 PM	patient 13	patient 7
5:00 PM-6:00 PM	patient 14	patient 14
6:00 PM-7:00 PM	patient 15	patient 15
7:00 PM-8:00 PM	patient 16	patient 10
8:00 PM-9:00 PM	patient 17	patient 13
9:00 PM-10:00 PM	patient 18	patient 18
10:00 PM-11:00 PM	patient 19	patient 19
11:00 PM-12:00 AM	patient 20	patient 20

Fig. 4 Sample schedule and resulting penalties

columns 'New order' and 'New priority' shows the newly arranged schedule for the patients and their corresponding priority levels. 'Timeslots moved' column gives the difference of the patient's new order from their original order of request. For example, Patient Number 4 is scheduled to be in time slot number 11 but was originally the 4th patient by order of requests. This means that Patient Number 4 was moved 7 time slots from the original order.

Figure 4 shows how the results from Model 2 can be operationalized given the assumption that one session is equivalent to one hour. The "Original" column represents the order by which patients call in for the appointment. Meanwhile, the "Base Run" column reflects the new schedule that takes into account both priority and order of appointment request. The penalties from both are also shown in Fig. 4. The resulting schedule from the model was able to obtain a lower schedule rating than the original order of the patients' requests. When comparing the components of the schedule rating, the model was able to minimize the total penalty or weight contribution from the priority levels of the patients. Although it incurred penalty from rearranging the original order, the overall schedule rating remained to be lower than the level obtained prior to using the model.

5 Conclusions

The total patient prediction model, referred to as Model 1, was developed in an attempt to consider the most unpredictable facets of patient scheduling in the form of lates, no-shows, and emergency patient arrivals. Unlike other models, which focused either on lates and no-shows or emergency patient arrivals exclusively, this model takes into account all of these facets when generating expected patient totals. The patient appointment scheduling model, referred to as Model 2, was created in order to address the scheduling aspect of the issue hospital systems face. The model sought to balance the decision between addressing the needs of high priority patients as soon as possible while at the same time honoring the initial appointment schedules made by patients.

The neighborhood search heuristic is then used to develop decision making criteria when uncertainties are realized. For the occurrence of no show patients, the results from the scenario analysis shows that moving the next immediate patient to the no show's supposed idle time provides a slightly lower schedule rating, as opposed to moving the next patient with the highest priority level. Therefore, it is recommended that when there is a no show, the scheduler will look at the base schedule and call the immediate next patient to the no show time slot.

References

1. Diagnostic Demand and Capacity Programme. Diagnostic demand and capacity report. South Yorkshire and Bassetlaw Integrated Care System (2019). https://www.healthandcaretogethersyb.co.uk
2. Van Sambeek, J., Joustra, P., Das, S., Bakker, P., Maas, M.: Reducing MRI access times by tackling the appointment-scheduling strategy. BMJ Qual. Saf. **20**(12), 1075–1080 (2011)
3. Burdett, R., Kozan, E.: An integrated approach for scheduling health care activities in a hospital. Eur. J. Oper. Res. **264**(2), 756–773 (2018)
4. Gupta, D., Denton, B.: Appointment scheduling in health care: challenges and opportunities. IIE Trans. **40**(9), 800–819 (2008)
5. Mahmoudzadeh, H., Shalamzari, A., Abouee-Mehrizi, H.: Robust multi-class multi-period patient scheduling with wait time targets. Oper. Res. Health Care **25** (2020). https://doi.org/10.1016/j.orhc.2020.100254
6. Hoseini, B., Cai, W., Abdel-Malek, L.: A carve-out model for primary care appointment scheduling with same-day requests and no-shows. Oper. Res. Health Care **16**, 41–58 (2018)
7. Deceuninck, M., Vuyst, S.D., Fiems, D.: An efficient control variate method for appointment scheduling with patient unpunctuality. Simul. Model. Pract. Theor. **90**, 116–129 (2019)
8. AlRowaili, M., Ahmed, A., Areabi, H.: Factors associated with No-Shows and rescheduling MRI appointments. BMC Health Serv. Res. **16**(1) (2016). https://doi.org/10.1186/s12913-016-1927-z

Discrete Event Simulation for Pharmaceutical Supply Chain Analysis in India

Rohit Sindhwani and Venkataramanaiah Saddikuti

Abstract Around the globe, pharmaceutical companies are increasingly becoming connected, internally as well as externally. Data and information is significantly important to deliver better quality healthcare. A large number of pharmaceutical business running legacy systems will not survive without investing in technology, data systems and innovative models. Hence, innovative business models warrants better product mix with complex generics and specialty products. We identify how supply chain management systems help in coordinating the planning and operations of emerging business models in pharmaceutical sector though network optimization. Since pharmaceutical products are for human consumption, complete control over distribution chain is mandatory. Hence, we focus on modern supply chain having traceable, secure, globalized and compliant logistics for the pharmaceutical sector. We simulate the pharmaceutical supply chain using digital twins. We utilize discrete event simulation methodology using anylogistix simulation and optimization toolkit. The case study in this work is the Jan Aarogya (peoples' health) scheme in India, which is built on the modern supply chain paradigm of hub and spoke architecture. We find that digital twins can significantly improve the decision-making during large-scale healthcare disruptions. We simulate pharmaceutical supply chains during disruptions by focusing on system resilience. A system resilience is composed of component resilience, buffer inventory and time. We derive the impact of disruptions and build a simulation model for the located pharmaceutical hubs. Based on our simulation model for pharmaceutical supply chains, we show the impact of network utilization and proactive planning. Suitable recommendations for policy makers is also highlighted based on the analysis results.

Keywords Pharmaceutical supply chain · Digital twin · Innovation · Disruption · Simulation

R. Sindhwani (✉)
Operations Management, IMT Ghaziabad, Ghaziabad, India
e-mail: efpm06001@iiml.ac.in

V. Saddikuti
Indian Institute of Management, Lucknow, India
e-mail: svenkat@iiml.ac.in

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_79

1 Introduction

Indian pharmaceutical industry produces 20% of the world's medicines. India is a source of 60,000 generic brands and has a network of 3000 pharma companies with more than 10,500 manufacturing facilities [1]. However, high out of pocket (OOPE) expenditure, the price of patented drugs, regulatory issues and spurious medications have pressurized the healthcare payers. Increased regulatory scrutiny has led to delay in product approvals. In recent years, the margins for large formulation players dropped significantly due to pricing pressure and weak product launches. As the world embraces COVID-19 pandemic, it is expected that the margins will stay flat because of high raw material costs and high cost of exports in 2021.

There are a number of challenges, which the pharmaceuticals industry is facing currently. These are primarily because of low productivity, rising research and development costs and reduced proprietary products. While broad challenges are important, we will look into some of the challenges faced by Indian Pharmaceutical industry. India is the largest provider of generic drugs as it supplies across the globe. However, OOPE in 2015–16 for India was at 60.5%, which was much higher that the global average of 20%. The condition is similar for state level burden on health expenses. Bihar's OOPE were around 80%, while in UP, the OOPE was around 75% of health expenditure. The share of medicines is the biggest contributor to OOPE incurred by households (Table 1). Even though the country is a hub for generic medicines, we see that high expenditure on drugs (Pharmaceutical share in OOPE), patented drug pricing, and dependency on China, significantly hinders the reach of essential medicines to people.

Table 1 Healthcare expenditure in India (Source: National Health Accounts)

S. No.	Indicators	NHA 2013–14	NHA 2014–15	NHA 2015–16	NHA 2016–17
1	GDP (In crores)	11,272,764	12,433,749	13,764,037	15,362,386
2	THE per capita (current prices)	3638	3826	4116	4381
3	THE as % of GDP	4.0	3.9	3.8	3.8
4	CHE as % of THE	93.0	93.4	93.7	92.8
5	OOPE as % of THE	64.2	62.6	60.6	58.7
6	Pharmaceutical expenditure as % of CHE	35.7	37.9	35.4	36.8

Legend: THE = Total heath expenditure, CHE = Current health expenditure, OOPE = Out of pocket expenditure.

In 2004, WHO estimated that 649 million Indians lacked access to essential medicines. To bridge this gap, the Government of India launched an emerging business model. The Government launched Pradhan Mantri Bhartiya Janaushadhi Pariyojna (PMBJP) in 2008 to sell quality generic medicines and surgicals in India. The accessibility of drugs in India has improved through Jan Aushadhi (JA) stores and e-Pharmacies post 2015. While the medical spending is increasing year on year, this scheme is accessible and affordable for lower income groups. The aim of this manuscript is to understand how the emerging business models like PMBJP can develop in pharmaceutical supply chain (PSC) in India. More specifically, we understand how supply chain management practices can help in delivering N95 masks to low income groups. Secondly, we explore the role of innovative digital twins that help in mitigating impact of large-scale disruptions on PSCs. Long duration disruption is modeled using discrete event simulation considering three scenarios with different recovery policies. The study led to answering following research questions:

- How can network optimization help in SC mapping of emerging business models in pharmaceutical sector in India?
- How does digital twins simulate the disruption propagation in pharmaceutical supply chains in India?

The manuscript is structured as follows. Section 2 focus on literature review on simulation based supply chain modeling and disruption propagation in supply chain. Section 3 highlights the problem framework for network optimization for emerging business models in pharmaceutical sector. Section 4 introduces the simulation model on the supply of generic drugs in JA stores in India, and Sect. 5 highlights the results and insights of the presented work. Finally, in Sect. 6, conclusion and future scope of work is recommended.

2 Literature Review

The pandemic crisis has led to disruption propagations across the supply chains [2–4]. The pandemic disruption is a low frequency/high impact event; however, the disruption is more of a dynamic process with an unpredictable timeline. The disruption propagation or ripple effect also influence bullwhip effect due to uncertain demand and backlog accumulation [5]. To overcome such a super disruption, supply chains need to adapt to new normal by focusing on resilience, survivability and viability [6]. The objective of the supply chains focus on maintaining the services critical to the society during the disruption and not just individual firms' performance. A PSC plays a crucial role in building the restoration efforts in a radically changed environment. India's role in stabilizing the service level of PSC is important as the country is referred to as the medical basket of the world. Further, [7] reflected the importance of simulation modelling to make informed decisions during COVID-19. We address the literature of digital twins in supply chain disruption propagation/ripple effect modelling [3] explored the effects of disruption propagation on supply chain

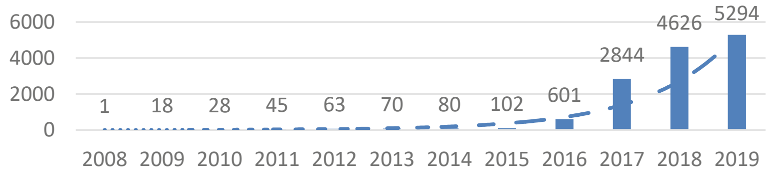


Fig. 1 Cumulative number of Jan Aushadhi stores over time

network. They identified that backup supply and flexible operations can help mitigate effects of forward and backward propagation respectively [4]. Studied the impact of epidemic outbreaks on SC performance and used anylogistix software to predict short-term and long-term impact on global SCs [8]. Analysed the impact of severe disruptions on the performance of healthcare supply chains so that there is no shortages in hospital wards. They identified backup supplier and lateral trans-shipment as the effective mitigation strategies for short term and long-term disruptions [9]. Found that disruption driven changes in supply chain behaviour causes backlog and delayed orders referred to as “disruption tails” [10, 11] provided important insights of COVID-19 on food supply chains using simulation. All these studies simulate to model the ripple effects in supply chains and consider supply and demand disruptions to analysis SC performance.

3 Problem Framework

3.1 Problem Statement

The main objective of JA scheme is to provide affordable, generic and high quality medicines through dedicated stores. There has been an exponential growth in number of JA store in the last decade (Fig. 1). The wholesale distributors serve the JA stores in the states. The JA scheme is built on the modern supply chain paradigm with hub and spoke architecture. We consider the requirement of N95 masks during the pandemic disruption. The network optimization model optimize the number of distribution centres and backup factories. The objective is to maximize the profit earned by the PSC of the JA scheme.

3.2 Notations and Decision Variables

The parameters and decision variables used in the model formulation are listed in Tables 2 and 3 respectively. Since the supply chain is built on the hub and spoke architecture, we consider the relevant central distribution centres (CDCs) as prospective

Table 2 Parameters and subscripts for the model formulation

Notations	Definition
i	Index for source nodes, $i \epsilon N$
j	Index for destination nodes, $j \epsilon N$
k	Index for hub connected to i, $k \epsilon H$
m	Index for hub connected to j, $m \epsilon H$
α	Discount factor for transshipment (hub to hub), $(k \rightarrow m)$
χ	Discount factor for collection (source to hub), $(i \rightarrow k)$
δ	Discount factor for distribution (source to hub), $(m \rightarrow j)$
N	Set of all nodes
H	Set of all hubs, $H \subseteq N$
W_{ij}	Flow from source i to destination j
d_{ij}	Cost of travelling from node i to node j
d_{ijkm}	Cost of traversing source i to destination j; $d_{ijkm} = \chi d_{ik} + \alpha d_{km} + \delta d_{mj}$
F_k	Fixed cost of establishing CDC at hub k
Γ_k	Capacity of hub k

Table 3 Decision variables for the model formulation

Notations	Definition
X_{ijkm}	Fraction of flow from source i to destination j, through hubs k and m
z_k	1 if pharmaceutical hub k remains open, 0 otherwise

hubs and model the hub location problem. The transportation from one CDC to other is through full truck load (FTL). Hence, we get economies of scale and a discount factor is attached to such transportation.

3.3 Mathematical Formulation

We model the Capacitated Multiple Allocation Hub Location Problem (CMAHLP) to locate the prospective CDC [12, 13]. We integrate the model with the PSC by discounting the trans-shipment costs between the CDCs. Thus, we find the location of the prospective distribution centre. The location for backup factories are found from standard transportation problem. The professional solver CPLEX is used in anylogistix to solve the supply chain model. Once we solve the network optimization model to find prospective supply chain configuration, we generate useful what-if analysis by simulating the digital twin. We incorporate disruption scenario events in our simulation.

$$\min_{X,z} G(X, z) = \sum_{i\in N}\sum_{j\in N}\sum_{k\in H}\sum_{m\in H} W_{ij} d_{ijkm} X_{ijkm} + \sum_{k} F_k z_k \tag{1}$$

$$s.t. \quad \sum_{k\in H}\sum_{m\in H} X_{ijkm} = 1 \quad \forall i, j \in N \tag{2}$$

$$\sum_{m\in H} X_{ijkm} + \sum_{m\in H/\{k\}} X_{ijmk} \le z_k \quad \forall i, j \in N, k \in H \tag{3}$$

$$\sum_{i\in N}\sum_{j\in N}\sum_{m\in H} X_{ijkm} \le \Gamma_k z_k \quad \forall k \in H \tag{4}$$

$$X_{ijkm} \ge 0 \quad \forall i, j \in N, k, m \in H \tag{5}$$

$$z_k \in \{0, 1\} \quad \forall k \in H \tag{6}$$

The objective function (1) minimizes the transportation and hub location cost. Constraint (2) states that demand between sources i to destination j has to be routed via hubs k and m. Constraint (3) is a facet defining constraint set which ensures that flow is sent via open hubs only. Constraint (4) is the capacity restriction of the open hub, Constraint (5) is the non-negativity constraint on the fractional flows and Constraint (6) is the binary constraint of the hub location.

4 Network Optimization and Simulation Model

4.1 Network Optimization

We utilize discrete event simulation methodology using anylogistix simulation and optimization toolkit. In particular, we consider the PSC for JA scheme in India. The supply chain is made up of four echelons: two suppliers, three factories, four central distribution centre (CDC) and thirty-seven wholesale drug distributors. The input data on which the study is based is presented in Table 4 (1 m^3 equivalent to 1000 pcs).

In the centralized configuration, the suppliers ship their products directly to CDCs. The CDC managed by BPPI (Bureau of Pharmaceutical Policy in India) then breaks down the shipment to smaller lots and directly ships to the JA Kendras (wholesale drug distributors in our case). There are two external suppliers (Venus and Magnum) for N95 surgical mask in the centralised configuration. Thus, the supply chain becomes more integrated and the flow becomes seamless.

In the decentralised case, we model three echelon supply chain where we have three internal backup factories of Indian drugs and pharmaceuticals limited (IDPL). These factories can be useful for production of N95 masks during surge in demand

Table 4 Input data for network optimization

S. No.	Parameter	Value
1	Demand for N95 masks in the markets per day (normally distributed)	1% of state's population, 2% during 2^{nd} wave (1 April–31 May)
2	Transportation distances and time in between supply chain facilities	Actual routes
3	Inventory carrying cost at CDC in $ per day, per m^3	10
4	Fixed facility costs, $ per day	5000
5	Transportation cost in $ per km, per m^3	0.01
6	Discount factor for transportation between CDC	0.5
7	Inbound processing cost at CDC, in $, per pcs	0.15
8	Outbound processing cost at CDC, in $, per pcs	0.1
9	Maximum CDC storage capacity, in pcs per day	300,000
10	Maximum production capacity at backup factory, in pcs per day	200,000
11	Production cost at backup factory, per pc in $	0.15
12	Cost of purchasing from primary suppliers, per pc in $	0.23
13	Unit Price, in $ per pc in $	0.34
14	Penalty for overutilization of production capacity, in $	100,000

or during disruptions. While the factories transport manufactured masks to CDCs, the CDC builds up inventory. The external suppliers only manufacture N95 masks. Figure 2 shows the distribution of facilities and clients and the optimized supply chain using network optimization.

4.2 Simulation Model

The assumptions for the experiments for digital twin simulation are presented below: The optimized supply chain from network optimization is considered for simulation experiment. A one-year period (Jan 1, 2021–Dec 30, 2021) is considered. The experimental part comprise the following disruption and reconfiguration scenarios (Table 5). The disruption scenarios are based on the increased number of cases in India during the second wave and lockdown restrictions. Scenario 1 is based on the increased demand of N95 masks during 2^{nd} wave. Scenario 2 looks into disruption

Fig. 2 Downstream JAK supply chain for N95 masks in India and optimized network using anylogistix

Table 5 Disruptions and reconfiguration scenarios

Scenario	Disruption	Recovery
1	Normal scenario (increased demand during 2nd wave)	Normal demand after 31 May
2	Full disruption of CDC 1 (Gurugram) on 1 April together with increased demand	50% CDC 1 capacity recovery on 1 May and full recovery on 30 June
3	Full disruption at Supplier 1 (Magnum) on 1 April together with increased demand	Full supplier recovery on 30 June

at CDC in Gurugram due to lockdown in Delhi-NCR. Scenario 3 is based on the disruption at Supplier in Maharashtra due to lockdown.

5 Results and Discussions

The optimal solution to our network optimization problem is to open CDCs in Gurugram and Benguluru, all open all the three backup factories. The supply chain design post network optimization (Fig. 2), after 10 iterations has an overall profit of 31,103,136 USD. The key performance indicators that we look into are financial performance, operational performance, and customer performance in normal state and disrupted state.

5.1 Financial Performance

As per the financial performance, total costs are 185.89 million, 99.29 million and 182.73 million USD in the three scenarios scenario respectively (Fig. 3). The costs reduce due to facility disruptions; however, the revenue and profit also decrease respectively. The supply chain for Scenario 2 faces loss of 1.17 million USD due to disruption of trans-shipment CDC hub.

5.2 Operational Performance

As can be seen in Fig. 4, there was a buildup of available inventory during the disrupted period for Scenario 2. There was a severe shortage during the disruption of supplier in Scenario 3. As far as lead time is concerned, we see that the lead time for the three scenarios remain around 24 days.

5.3 Customer Performance

As can be seen in Table 6, the service level reduce in Scenario 2 once the disruption occurs and the backlog keeps building in future. Also, the demand placed by site drops significantly in Scenario 2 since only one CDC is fully functional during disruption. The ELT service level for N95 masks deteriorates in Scenario 3 (indication of impact of ripple effect). The bullwhip effect is higher in Scenario 3 since the disruption is at the upstream, and thus impacts more SC entities.

The simulation model investigated the impact of large-scale disruption and the effects with backup factories in place. For each scenario, financial, operational and customer performance is carried out. We understand that system resilience and flexibility need to built in the model as the PSC is important for the overall well-being of the individual. The challenges in PSCs are reducing counterfeit drugs and maintaining time and temperature controlled transport. If the inventory keeps building ups, obsolescence level will increase resulting in wastages. The distribution challenges that accompany cold chain operations results in potentially life threatening consequences.

6 Conclusion

In this paper, we applied network optimization to the emerging business models in pharmaceutical sector. We analysed the impact of large-scale disruptions on PSCs. Combining the emerging business models with digital twins using anylogistix, we build some novel insights through performance indicators.

We find that the broad challenges that impact the pharmaceutical industry is demand forecasting, price fluctuation assessment and disruption management. Demand forecasting techniques require that pharmaceutical goods are produced and delivered seamlessly in time. This goal can be accomplished with the use of healthcare

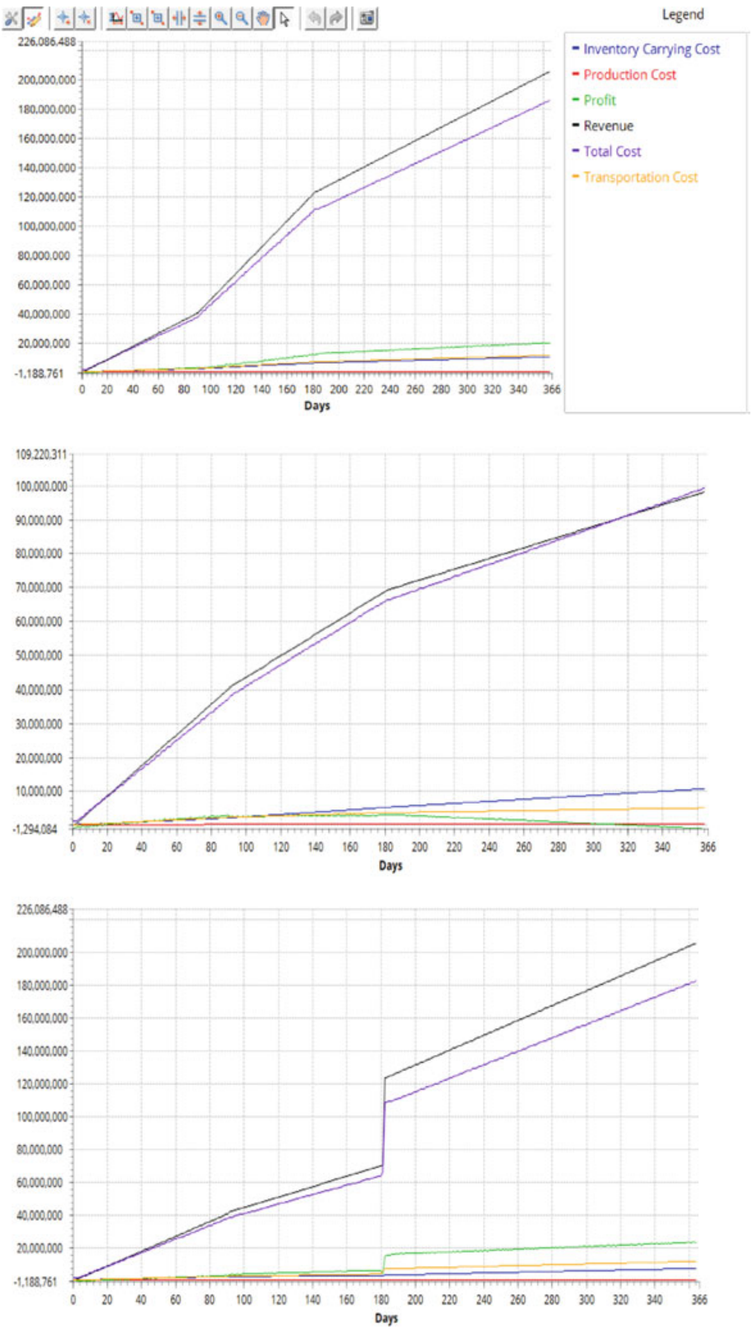


Fig. 3 Financial performance of the PSC in three scenarios

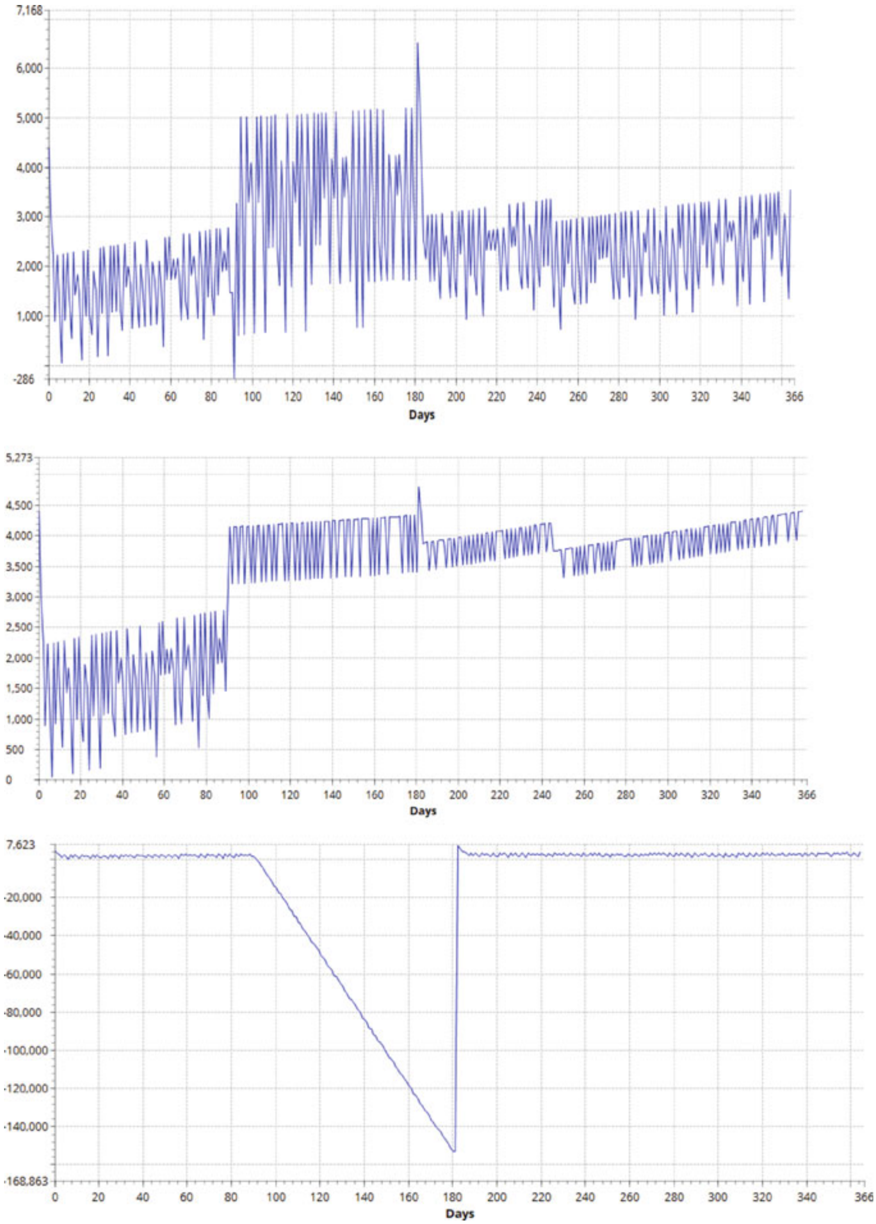


Fig. 4 Available inventory including backlog

Table 6 Key performance indicators (KPIs) for customer performance

KPI	Scenario 1	Scenario 2	Scenario 3
Demand Placed (Orders) by Site	393.0	277.0	358.0
ELT Service Level by Orders	1.0	1.0	0.8965
Service Level by Orders	0.9965	0.5279	0.8435
Bullwhip Effect by Product	1.0191	0.9999	1.2757

analytics and data mining techniques. Secondly, analysing pricing structures for low-income group in India is another challenge in pharmaceutical industry. Thirdly, there is always a risk of disruption propagation in the PSC. We highlight that simulating the propagation of counterfeit drugs though online pharmacies and incorporating perishability can be an important area of research.

References

1. Ghosh, P.: Novel active pharmaceutical ingredients from India: the actors-part-I. MGM J. Med. Sci. **8**, 73 (2021). https://doi.org/10.4103/mgmj.mgmj_12_21
2. Ivanov, D., Dolgui, A.: OR-methods for coping with the ripple effect in supply chains during COVID-19 pandemic: managerial insights and research implications. Int. J. Prod. Econ. **232**, 107921 (2021). https://doi.org/10.1016/j.ijpe.2020.107921
3. Li, Y., Chen, K., Collignon, S., Ivanov, D.: Ripple effect in the supply chain network: forward and backward disruption propagation, network health and firm vulnerability. Eur. J. Oper. Res. **291**, 1117–1131 (2021). https://doi.org/10.1016/j.ejor.2020.09.053
4. Ivanov, D.: Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis on the coronavirus outbreak. Transp. Res. Part E. **136**, 101922 (2020). https://doi.org/10.1016/j.tre.2020.101922
5. Dolgui, A., Ivanov, D., Rozhkov, M.: Does the ripple effect influence the bullwhip effect? An integrated analysis of structural and operational dynamics in the supply chain†. Int. J. Prod. Res. **58**, 1285–1301 (2020). https://doi.org/10.1080/00207543.2019.1627438
6. Ivanov, D., Dolgui, A.: Viability of intertwined supply networks : extending the supply chain resilience angles towards survivability: a position paper motivated by COVID-19 outbreak. Int. J. Prod. Res. 58, 1–12 (2020). https://doi.org/10.1080/00207543.2020.1750727
7. Currie, C.S.M., Fowler, J.W., Kotiadis, K., Monks, T., Onggo, B.S., Robertson, D.A., Tako, A.A.: How simulation modelling can help reduce the impact of COVID-19. J. Simul. **14**, 83–97 (2020). https://doi.org/10.1080/17477778.2020.1751570
8. Aldrighetti, R.: Healthcare supply chain simulation with disruption considerations: a case study from Northern Italy. Glob. J. Flex. Syst. Manag. **20**, 81–102 (2019). https://doi.org/10.1007/s40171-019-00223-8
9. Ivanov, D.: Disruption tails and revival policies: a simulation analysis of supply chain design and production-ordering systems in the recovery and post-disruption periods. Comput. Ind. Eng. (2018). https://doi.org/10.1016/j.cie.2018.10.043
10. Burgos, D., Ivanov, D.: Food retail supply chain resilience and the COVID-19 pandemic: a digital twin-based impact analysis and improvement directions. Transp. Res. Part E. **152**, 102412 (2021). https://doi.org/10.1016/j.tre.2021.102412

11. Singh, S., Kumar, R., Panchal, R., Tiwari, M.K.: Impact of COVID-19 on logistics systems and disruptions in food supply chain. Int. J. Prod. Res. **59**, 1993–2008 (2021). https://doi.org/10.1080/00207543.2020.1792000
12. O'Kelly, M.E.: The location of interacting hub facilities. Transp. Sci. **20**(2), 92–106 (1986). https://doi.org/10.1287/trsc.20.2.92
13. Ebery, J., Krishnamoorthy, M., Ernst, A., Boland, N.: The capacitated multiple allocation hub location problem: formulations and algorithms. Eur. J. Oper. Res. **120**(3), 614–631 (2000). https://doi.org/10.1016/S0377-2217(98)00395-6

Dual Resource Scheduling in Trauma Care Centre with Time Varying Patient Demand

Venkataramanaiah Saddikuti and Pavan Kumar Gudavalleti

Abstract This paper presents an integrated approach for the dual resource (medical staff and emergency beds) scheduling at each Point of Care (PoC) under time-varying Patient arrival demand for trauma centers. Practical constraints, such as the number of beds available, the number of patients that can be treated simultaneously at each Point of Care (PoC), and the signing on/off medical staff, are considered. The objective is to reduce the Patient's waiting time before entering into Operation Theatre/discharged since the Patient's arrival to the hospital. The mixed-integer linear programming (MILP) technique is used to solve the resulting multi-objective to deliver medical staff scheduling (i.e., signing on and off times of medical staff) and bed requirements simultaneously. The reduction of waiting times with this proposed algorithm is analyzed. A case study is conducted based on real-life data from a trauma center in India. The proposed approach is compared with the practical approach being followed in the trauma center. This comparison shows the proposed approach's effectiveness, suggesting scheduling medical staff based on historic patient arrival patterns will help reduce Patient's waiting times before entering Operation Theatre since the patient arrival. There is an average improvement in total process times for patients by 30%, with enhancements considered at a critical PoC.

Keywords Emergency care · Capacity planning in trauma care center · Medical resources optimization · Point of care

1 Introduction

Timely support of emergency care is crucial for the recovery of the Patient in need of it. During workdays or weekends, the hourly rate of patient arrival varies and

V. Saddikuti (✉) · P. K. Gudavalleti
Indian Institute of Management Lucknow, Lucknow, India
e-mail: svenkat@iiml.ac.in

P. K. Gudavalleti
e-mail: efpm07002@iiml.ac.in

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_80

can be estimated from historical data. As patient arrival rates increase and beds become unavailable and the attention of medical personnel becomes saturated. Thus, the planning process demands great attention because the recovery of the Patient is strongly influenced by qualified and timely attention.

As shown in Fig. 1, the different stages of trauma treatment include external factors (ambulance services), internal factors (primary treatment and theatre of operations), and post-operative treatment. This paper focuses only on how to optimize primary treatment until the Patient arrives in an operating Theatre. The primary treatment section is considered in three broad categories: Human Resources, Physical Resources and Process—Organization and administration (Mock C, 2004). Traditionally, medical staff planning (Human Resources) and bed availability (Physical Resources) are planned only once, in the initial stages of setting up the trauma care center.

In general, failure to update the planning process can cause the trauma center to operate in a sub-optimal manner. Planning determines the number of beds and hours of employment of specialized, general, nursing and paramedical staff to meet the demand for incoming patients within the constraints of the infrastructure. The bed requirements at each PoC can be determined based on the historical patient arrival rates, based on which the feasible schedule of medical staff could be obtained for a given interval. We focus on integrating medical staff scheduling and bed requirement plans for a typical trauma care center. In this optimization problem, detailed explanations regarding the assumptions are given in Sect. 3.2. In this paper, an optimization model is proposed for the medical staff scheduling and bed circulation planning for a trauma care center, with an objective to minimize the waiting times of the patients before entering the operation theatre. A mixed-integer linear programming (MILP) model is formulated, where the time-varying patient demand, the medical staff scheduling, the connection of bed services at each Point of Care are included in the model formulation. The rest of this document is organized along the following lines. Section 2 provides a literature review on the effect of waiting times of the Patient due to their arrival rate. Section 3 provides a detailed problem statement and the assumptions for formulating the model. In Sect. 4, the optimization problem for patient demand-oriented medical staff scheduling is formulated. In Sect. 5, a case study in a local trauma care center at Lucknow in India and the implications of the model are discussed.

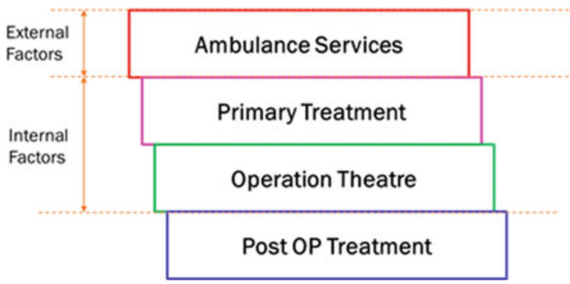


Fig. 1 Trauma Care delivery process

2 Literature Review

The purpose of a trauma care center is to stabilize patients who have a life-threatening or limb-threatening injury or illness and it focuses on the provision of immediate medical interventions [1]. After an accident or incident, the Patient is transported to the hospital by ambulance services. Pre-hospital emergency care is provided when the Patient needs life support until reaching the trauma center [2]. The trauma care center faces significant challenges in delivering high-quality, timely care to patients, given the increasing number of patients and limited hospital resources [3]. Several studies indicated that patient crowding contributes to reduced quality of patient care [4], treatment delays [5, 6], and worsens adherence with medical authority guidelines [7].

Modeling Approaches for Emergency department patient flow are classified into formula-based, regression-based, time-series analysis, queuing theory and Discrete Event Simulation [8]. The formula-based methodology employs past flow performance based on the number of staff and beds. Typical assessment tools include Emergency Department Work Index [9], Work Score [10, 11], Early Warning System and Demand Value [12]. They are straightforward and are used to divert ambulance services but do not have scalability and do not employ optimization techniques. The regression, a statistical technique uses the number of patients, determining patients' waiting times, staffing and information from other services, [13–15]. Due to inherent limitations in regression, for better handling, time-series based approaches are used for patient arrivals, bed occupancy, length of stay etc. [16–19]. The time-series method fails to capture the level of short-term variability. Queuing models help in optimal staffing and patient prioritization. These models generate relations between capacity, waiting times and treating process times. These models approximate patient arrival rates (Poisson/random) and flow patterns [20, 21]. They work on targeting discharging of patients within a set target. Discrete-event Simulation mime the models above with a graphical user interface helps in "What-If" analysis but cannot be used for "best-fit" models [8].

In this paper, we present an optimization model, with an objective to minimize the waiting times of the patients before entering the operation theatre by adaptively scheduling the medical staff based on patient arrival pattern. This would fetch better patient treatment satisfactorily and facilitates identification of bottleneck for future expansion plans. The average patient arrival rate at a trauma care center is around 15–20 per day in cities of India [22–25].

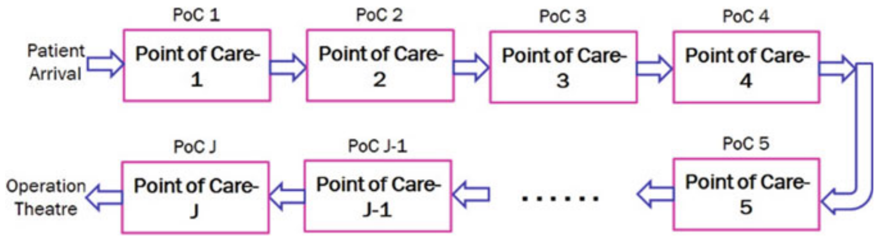


Fig. 2 Steps in emergency care delivery at the trauma center considered in the study

3 Problem Statement and Formulation

3.1 Problem Statement

We consider a trauma care center with the steps given in Fig. 2, where the patients can enter the trauma center from the Point of Care, PoC 1. Intended care is delivered to the patients at each Point of Care. The total number of PoCs is J. The patient treatment direction is from PoC 1 to J. Re-examination, and patient review at each Point Care is considered an additional PoC in the basic model. In a practical scenario, re-examination will be carried out at the earlier PoC. The number of patients treated simultaneously at each PoC depends on patient arrival and resource availability at the given PoC.

When considering medical staff scheduling problems for a trauma center, the inter-arrival time of patients changes several times; the incoming patient arrival rate in the peak hours may be significantly higher than that in the off-peak hours. Typical schedules for offering services to 4 patients in an hour, each arriving at an interval of 15 min, are shown in Fig. 3. This trauma center is assumed to have 9 PoCs before the Patient enters the Operation theatre. We propose to obtain Medical staff scheduling with the consideration of the number of simultaneous services to the Patient that can be offered at each Point of Care and patient demand to minimize the patients' total waiting time before entering into Operation Theatre.

3.2 Assumptions

The detailed descriptions of the assumptions to obtain Patient demand-oriented medical staff scheduling are described below:

(1) Each point where a Patient is observed, like primary care, fee counter, etc., is termed as a Point of Care (PoC);
(2) Patient in observation at a PoC cannot be set aside to treat another patient;
(3) Each bed in a PoC can accommodate only one Patient at a time;
(4) The available medical staff can be scheduled during the whole operating period;

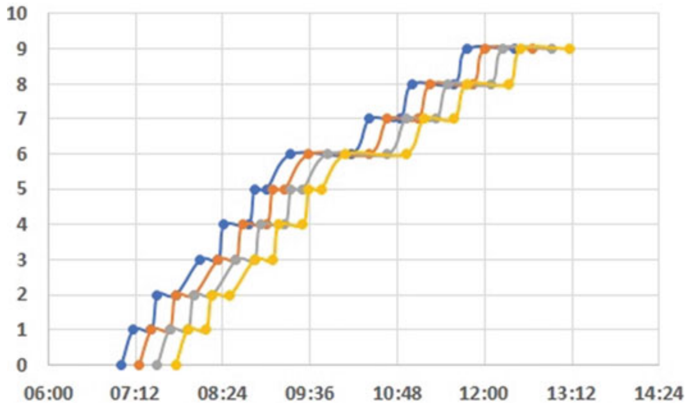


Fig. 3 Example of a patient treatment plan since arrival till entering operation theatre

(5) The medical staff can be scheduled for a continuous minimum and maximum working hours;
(6) The number of medical staff is the effective number after considering leave reserve and rest giver;
(7) There is a restriction on the maximum number of hours that medical staff can be called for working in a week.

4 Mathematical Formulations

The mathematical model is presented in this section to optimize the medical staff scheduling and bed requirement at each Point of Care. Initially, the notations and decision variables are discussed. Then, the constraints and objective functions of the patient arrival rate-based medical staff scheduling is formulated.

4.1 Notations and Decision Variables

The parameters and subscripts used in the model formulation are listed in Tables 1 and 2 lists the decision variables. Since the model is built based on patient arrival rates, the departure and arrival times at Point of Cares (PoCs), the transit times, waiting times and process times at each PoC are considered essential elements. Moreover, the availability of active beds at each PoC is a necessary aspect of planning. The considered possibilities include (1) whether the bed needs to activated or not, (2) whether the bed is to be deactivated or not immediately after the service to a patient, and (3) whether the same bed is to be used to serve the other Patient arrived later or not.

Table 1 Parameters and subscripts for the model formulation

Notations	Definition
$\mathbb{I}$	Set of patients {1,2,3,….I}
i, i'	Index of patients $i, i' \in \mathbb{I}$
$\mathbb{J}$	Set of point of care (PoC) {1,2,3,….J}
j, j'	Index of point of care $j, j' \in \mathbb{J}$
k	Number of time intervals in the operating period, where $k \in [1, 2, 3, \ldots, \text{K}]$
tra_j	Transit time from PoC j to $j + 1,\ \text{j} \in \mathbb{J}/\{J\}$
$wt_{j,min}$	Minimal Waiting time for the process to start after arrival at PoC $j \in \mathbb{J}$
$wt_{j,max}$	Maximal Waiting time for the process to start after arrival at PoC $j \in \mathbb{J}$
$\tau_{i,j}$	Process time of patient i at point of care $j \in \mathbb{J}$
$\tau_{j,min}$	Minimum process time at PoC $j \in \mathbb{J}$
$\tau_{j,max}$	Maximum process time at PoC $j \in \mathbb{J}$
t_{start}	Start time of the trauma center services, $t_{start} = t_0$
t_{end}	End time of the trauma center services, $t_{end} = t_K$
p_k	Number of PATIENTS Arriving in time interval $\left[t_{k-1}, t_k\right)$ where $k \in [1, 2, 3, \ldots, \text{K}]$
λ_k	Patient arrival rates in time interval $\left[t_{k-1}, t_k\right)$ where $k \in [1, 2, 3, \ldots, \text{K}]$
$N_{bed,j}$	Maximum number of beds available at Point of Care $j \in \mathbb{J}$
M	Large number

Table 2 Decision variables for the model formulation

Notations	Definition
$pstart_{i,j}$	Process start time of patient $i \in \mathbb{I}$ at point of care j $\in \mathbb{J}$
$pend_{i,j}$	Process end time of patient $i \in \mathbb{I}$ at point of care j $\in \mathbb{J}$
$a_{i,j}$	Arrival time of patient $i \in \mathbb{I}$ at point of care j $\in \mathbb{J}$
$wt_{i,j}$	Waiting time of patient $i \in \mathbb{I}$ at point of care j $\in \mathbb{J}$
$\chi_{i,k}$	0–1 binary variable, $\chi_{i,k} = 1$, if the patient $i \in \mathbb{I}$ at point of care 1 in the interval $\left[t_{k-1}, t_k\right)$, Otherwise $\chi_{i,k} = 0$
$\xi_{i,j}$	0–1 binary variable, $\xi_{i,j} = 1$, if the new bed service is initiated to offer service to Patient $i \in \mathbb{I}$ at Point of Care j $\in \mathbb{J}$, Otherwise $\xi_{i,j} = 0$
$\delta_{i,j}$	0–1 binary variable, $\delta_{i,j} = 1$, if the bed service is withdrawn after service to Patient $i \in \mathbb{I}$ at Point of Care j $\in \mathbb{J}$, Otherwise $\delta_{i,j} = 0$
$\beta_{i,i',j}$	0–1 binary variable, $\beta_{i,i',j} = 1$, if the bed service at Point of Care j $\in \mathbb{J}$ is connected to offer service to Patient $i \in \mathbb{I}$ and $i' \in \mathbb{I}$, Otherwise $\beta_{i,i',j} = 0$

4.2 Constraints

In this subsection, constraints are discussed and formulated to generate a feasible schedule of the activation or deactivation of beds at each PoC based on the historic patient arrival rate.

Starting Constraints

The patient arrival rate has these scenarios as illustrated in Fig. 4, where the arrival rates of two successive patients are estimated to be the same in a given time interval, i.e., the patient arrival rate is constant between the arrival of Patient i and i−1 within the interval.

Assuming that only arrivals are possible in the same interval, then the arrival time of patent i at PoC 1 can be indicated as the following constraint:

$$a_{i,1} \leq t_{k,start} + tra_{i,1} + (m-1).\lambda_k \quad \forall i \epsilon I, k \epsilon K, m \in p_k \tag{1}$$

Arrival Constraints of Patient at each Point of Care.

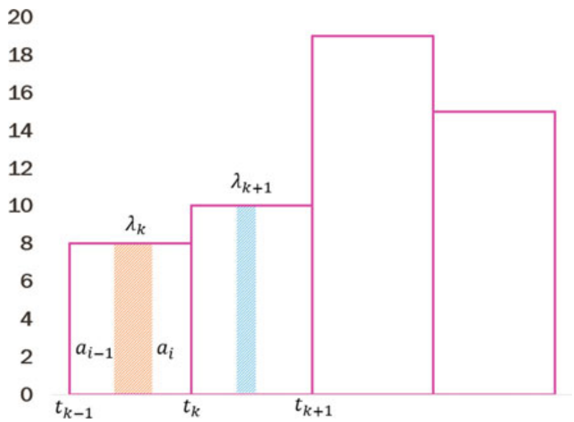
The various time processes involved in patient movement from one Point of Care to another are shown in Fig. 5. The figure illustrates. Process end time, Arrival time and Process start time are flow variables, whereas the Process time, Transit time and Waiting time are the interval time variables.

The arrival time of patient i at Point of Care j is equivalent to the sum of the process end time at Point of Care $j-1$ and the transit time from the Point of Care $j-1$ to j, except for the first Point of Care. The arrival time of patient i at Point of Care j should satisfy the following constraint:

$$a_{i,j} = pend_{i,j-1} + tra_{i,j} \, \forall j \in \mathbb{J}/1 \tag{2}$$

Process Start Constraints of Patient at each Point of Care.

Fig. 4 Scenarios for calculation of Patient arrival



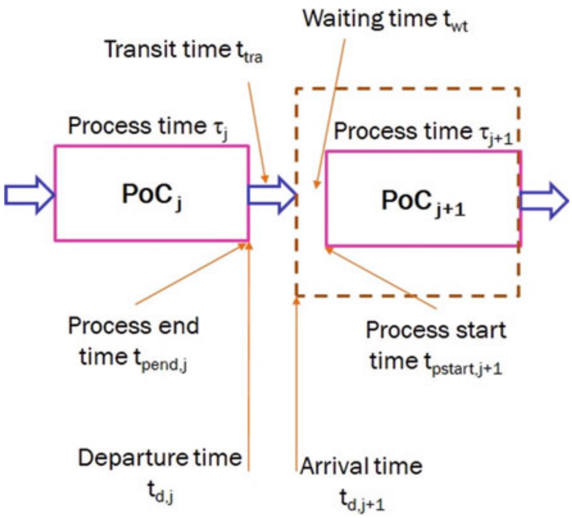


Fig. 5 Various times involved in patient movement from one point of care to another

The process start time of patient i at Point of Care j is equivalent to the sum of the arrival time and the waiting time at Point of Care j. The process start time of patient i at Point of Care j should satisfy the following constraint:

$$pstart_{i,j} >= a_{i,j} + wt_{i,j} \tag{3}$$

Since a patient being offered service cannot be kept aside for providing service to another patient, according to Assumption (2), the Patient's order of services holds the same for all the Point of Cares if the beds are available.

$$pstart_{i,j} - pend_{i-k,j} >= 0 \forall i > N_{beds,j} \, and \quad k = \{1, \ldots i\} \tag{4}$$

Process End Constraints of Patient at each Point of Care.

The process end time of patient i at Point of Care j is equivalent to the sum of the process start time and the process time at Point of Care j. The process start time of patient i at Point of Care j should satisfy the following constraint:

$$pend_{i,j} = pstart_{i,j} + \tau_j \tag{5}$$

Process time Constraints.

The process time of Patient i at Point of Care j should satisfy the following constraint:

$$\tau_{j,min} \leq \tau_{i,j} \leq \tau_{j,max} \tag{6}$$

Transit time Constraints.

The transit time of Patient i from Point of Care $j-1$ to j should satisfy the following constraint:

$$tra_{j,min} \leq tra_{i,j} \leq tra_{j,max} \tag{7}$$

Waiting time Constraints.

The Waiting time of Patient i from Point of Care j should satisfy the following constraint:

$$wt_{j,min} \leq wt_{i,j} \leq wt_{j,max} \tag{8}$$

Bed Constraints.

If services of patient i and i' utilize the same bed at Point of Care, then $\beta_{i,i',j} = 1$. However, if patient services i and i' does not utilize the same bed at Point of Care j, i.e., $\beta_{i,i',j} = 0$, then it means that they use different beds at Point of Care j. A large positive number M is introduced, and these constraints at PoC j are formulated as:

$$pend_{i',j} - a_{i,j} \geq wt_{j,min} - M\left(1 - \beta_{i,i',j}\right) \quad \forall i \epsilon \mathbb{I}/1, i' \epsilon \mathbb{I}, j \in \mathbb{J}/1 \tag{9}$$

$$pend_{i',j} - a_{i,j} \leq wt_{j,max} + M\left(1 - \beta_{i,i',j}\right) \quad \forall i \epsilon \mathbb{I}/1, i' \epsilon \mathbb{I}, j \in \mathbb{J}/1 \tag{10}$$

When $\beta_{i,i',j} = 0$, i.e., patient i and patient i' does not use the same bed at PoC j, the constraints (9) and (10) become $pend_{i,j} - a_{i',j} \geq wt_{j,min} - M$ and $pend_{i,j} - a_{i',j} \leq wt_{j,max} + M$ respectively. Due to the big M value, these values are satisfied. When $\beta_{i,i',j} = 1$, i.e., patient i and patient service i' use the same bed at PoC j, the constraints (9) and (10) become $pend_{i,j} - a_{i',j} \geq wt_{j,min}$ and $pend_{i,j} - a_{i',j} \leq wt_{j,max}$, which mean that the patients at PoC j, satisfy the minimal and maximal waiting time constraints.

The bed assigned to provide service to a patient at PoC j can be initiated or can be the same used by earlier Patients. It means that when a current patient is offered service at a bed, it can be withdrawn or used to perform service to another patient. Based on the definitions of the binary variables $\xi_{i,j}$ and $\beta_{i,i',j}$ in Table 3, if there is no existing activated bed can be used to service patient i', i.e., $\sum_{i \in \mathbb{I}} \beta_{i,i',j} = 0$, then new bed for offering service to the patient i' must be assigned, if feasible else the Patient shall wait. Thus, the following constraint is formulated:

$$\xi_{i',j} = 1 - \sum_{i \in \mathbb{I}} \beta_{i,i',j} \tag{11}$$

If the bed serving patient i' is not new, then there must exist only one bed to service patient i', i.e., $\sum_{i \in \mathbb{I}} \beta_{i,i',j} = 1$, since $\xi_{i',j}$ is a binary variable and its maximum value is equal to 1. Thus, the following constraint is formulated:

Table 3 Sample Example

Patient No	Arrival time (minutes)	POC1 Process end time (minutes)	POC2 Process end time (minutes)	POC3 Process end time (minutes)	POC4 Process end time (minutes)	Total Process time (minutes)
P1	0	20	70	100	132	132
P2	13	33	83	113	147	134
P3	17	43	93	123	162	145
P4	24	53	103	133	177	153
P5	36	63	113	143	192	156

$$\sum_{i \in \mathbb{I}} \beta_{i,i',j} \leq 1 \tag{12}$$

If the bed after offering service to patient i is not offering further service within the interval, i.e., $\sum_{i' \in \mathbb{I}} \beta_{i,i',j} = 0$, then the bed after serving patient i is withdrawn from the operation. Otherwise, the bed after serving patient i continues to offer service to another patient. Thus, the following constraint is formulated:

$$\delta_{i,j} = 1 - \sum_{i' \in \mathbb{I}} \beta_{i,i',j} \tag{13}$$

Each bed can perform service to only one Patient at a time. Thus, the following constraint is formulated:

$$\sum_{i' \in \mathbb{I}} \beta_{i,i',j} \leq 1 \tag{14}$$

Also, if the same bed offers services to two patients i and i' with $i < i'$, when $\beta_{i,i',j} = 1$, then the bed serving patient i must not be withdrawn, i.e., $\delta_{i,j} = 0$. But this bed shall also perform service to patient i'. Hence, the bed serving patient i' does not require the new bed activated, i.e., $\xi_{i',j} = 0$. Thus, the bed activation should satisfy the following constraints:

$$\delta_{i,j} + \xi_{i',j} \leq M\left(1 - \beta_{i,i',j}\right) \tag{15}$$

$$\delta_{i,j} + \xi_{i',j} \geq -M\left(1 - \beta_{i,i',j}\right) \tag{16}$$

When $\beta_{i,i',j} = 0$, the constraints are given in (15) and (16) will be satisfied due to the big M.

Practically, the number of beds available for the activation is limited and is denoted by $N_{bed,j}$. When $\xi_{i',j} = 1$, i.e., the bed offering service to patient i being newly assigned, we need to check whether there are still beds available at PoC j to perform service to Patient i. Thus, the difference between the total number of beds deactivated from utilization and the total number of beds under utilization should be less than or equal to the number of available beds at that Point of Care, j. This condition can be formulated as:

$$\sum_{m=1}^{i} \xi_{m,j} - \sum_{m=1}^{i-1} \delta_{m,j} \leq N_{bed,j} \tag{17}$$

where $\sum_{m=1}^{i} \xi_{m,j}$ and $\sum_{m=1}^{i-1} \delta_{m,j}$ are the total number of beds activated and deactivated until the departure of patient service i from PoC j.

Patient Demand Constraints

The Patient arrival rates can be written as:

$$\tilde{\lambda}(t) = \begin{cases} \lambda_1, & if t \varepsilon [t_0, t_1) \\ \lambda_2, & if t \varepsilon [t_1, t_2) \\ & \cdots \\ \lambda_K, & if t \varepsilon \left[t_{K-1}, t_K\right) \end{cases} \tag{18}$$

whereas

$$\lambda_k = \frac{\kappa_k}{(t_k - t_{k-1})} \tag{19}$$

where κ_k is the number of patients arriving in the interval $[t_{k-1}, t_k)$. The operating period $[t_{start}, t_{end}]$ is split into K time slots with the splitting time instants $t_1, t_2, \ldots, t_{K-2}, t_{K-1}$ where t_0 is t_{start} and t_K is t_{end}. The patient arrival rates are assumed as Poisson distributed within the interval.

4.3 Objective Function

An even bed activation and deactivation plan at each PoC between successive patients entering into Operation theatre can reduce total waiting time, based on the patient arrival pattern. The objective function is to minimize the variation between the consecutive end time of the patient treatment process at each Point of Care to reduce the waiting time of the patients. In an ideal case, the inter-arrival time of patients shall be the same as the inter-departure time of patients to the operation theatre. Thus, an objective function is formulated as follows:

$$Minimize \sum_{i \in \mathbb{I}/1} \left(a_{i,1} - a_{i-1,1}\right) - \left(pend_{i,J} - pend_{i-1,J}\right) \tag{20}$$

Table 4 Various time parameters in Minutes at each Point of Care

Point of Care	Transit time to next PoC (Min)	Transit time to next PoC (Max)	Waiting time (Min)	Waiting time (Max)	Process time (Min)	Process time (Max)	Capacity of Patients treatment at each POC
POC1: Stretcher	20	30	10	40	10	15	2
POC2: Primary check	10	20	20	70	10	30	2
POC3: Cash payment	5	10	15	45	5	20	3
POC4: Lab investigation	2	5	120	240	15	120	1
POC5: Ultrasound investigation	5	15	10	30	10	20	2
POC6: Re-examination	10	25	15	45	10	45	1
POC7: Surgery ward	5	15	40	180	15	25	1
POC8: Surgeon check	5	10	30	120	10	20	1
POC9: Speciality Ward	15	20	130	240	5	15	2
POC10: Specialist check	5	20	15	90	5	20	1
POC11: Operation theatre	30	45	720	4320	20	45	1

To explain the objective function, a sample example assuming a typical hospital with four PoCs is discussed in Table 4. Let the Patients are arriving hospital at arrival times of 0,13,17,24 and 36 min. The total process time of each Patient is 132, 134, 145, 153 and 156 respectively. It can be seen that the process time on each Patient is increasing, if the Patient inter-arrival time is reducing. Ideally, the process time for all the patients should be same i.e., the difference between the process times of Patient i and his/her predecessor Patient $i-1$ shall be Zero. Hence, the objective function is chosen to minimize the process times of two successive patients.

5 Case Study

The MILP programming is carried out through CPLEX (student version, for less than 1000 variables) and Matlab. The proposed model has been applied to a Trauma Center in Lucknow. The average number of cases seen daily is 110 and the average number of admissions through this center daily is 70 [1]. The center will function round the clock on all days throughout the year. There are 150 beds and ten operation theatres (five major and five minor). The layout of this trauma center is illustrated in Fig. 6. The detailed activities taken at each PoC are discussed in [2].

The transit, waiting and process time is shown in Table 4. Transit time and Process time are obtained from [2]. The minimum and maximum waiting times are obtained from the trauma center and represent the performance of the trauma center. The operating period of the trauma center is 24 h.

The patient arrivals are approximated in intervals of three hours based on the overall road accident rates in India [3]. The estimated patient arrivals are shown in Fig. 7. It can be noted that the peak hours are between 09:00 AM to 09:00 PM. Based on this historic arrival pattern, the number of facilities at each PoC can be put into operation at that time interval can be planned to minimize the patient waiting times.

If the capacity is doubled at each PoC, the percentage improvement is around 30%. See Table 5. The results of the first three Patient's travel are plotted in Fig. 8 as programmed in Matlab. Using the optimization model, it can be seen that most of

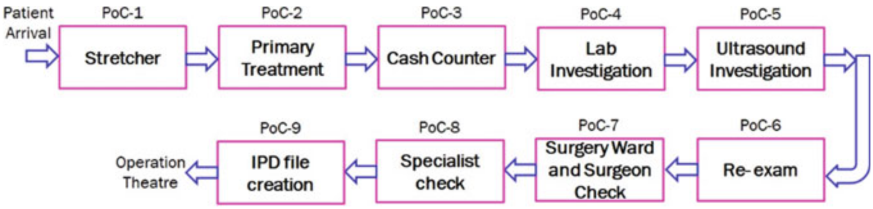


Fig. 6 The layout of Trauma center operations

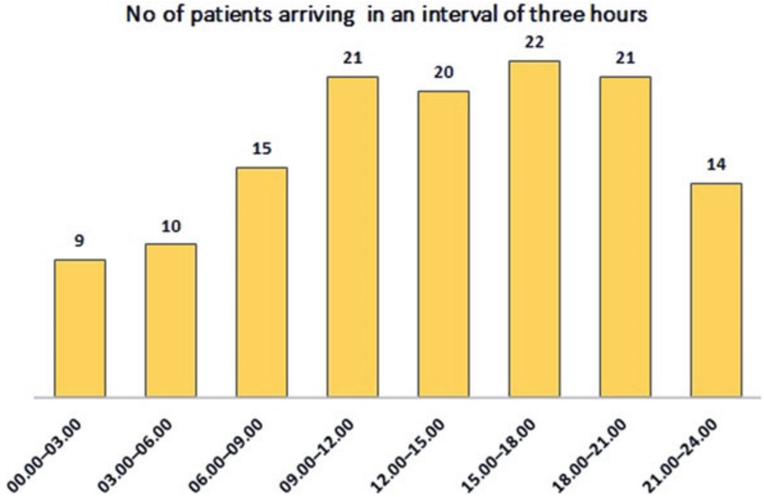


Fig. 7 Estimated Patient arrivals in three hours interval

the waiting time is available in PoC4. Now keeping bed capacity at all other PoCs as one and double the capacity at the PoC4 yielded almost similar results as shown in Table 6 with an only enhancement of capacity in PoC 4.

From the above, it can be seen that the performance of the hospital in terms of delivering care to the hospital can be optimized using this model. The ratio of beds activated to the staff requirement is constant. It can be obtained using multiplication

Table 5 Increase in the improvement of process time by doubling the capacity at each PoC

Patient No	Arrival time (minutes)	Process end time (minutes) till PoC4, when only one bed is operated in each PoC	Process end time (minutes) till PoC4, when two beds are operated in each PoC	Percentage improvement
P1	0	132	132	NA
P2	13	134	132	2%
P3	17	145	132	10%
P4	24	153	136	13%
P5	36	156	132	18%
P6	47	160	132	21%
P7	64	158	132	20%
P8	68	169	132	28%
P9	76	176	135	30%
P10	93	174	132	32%

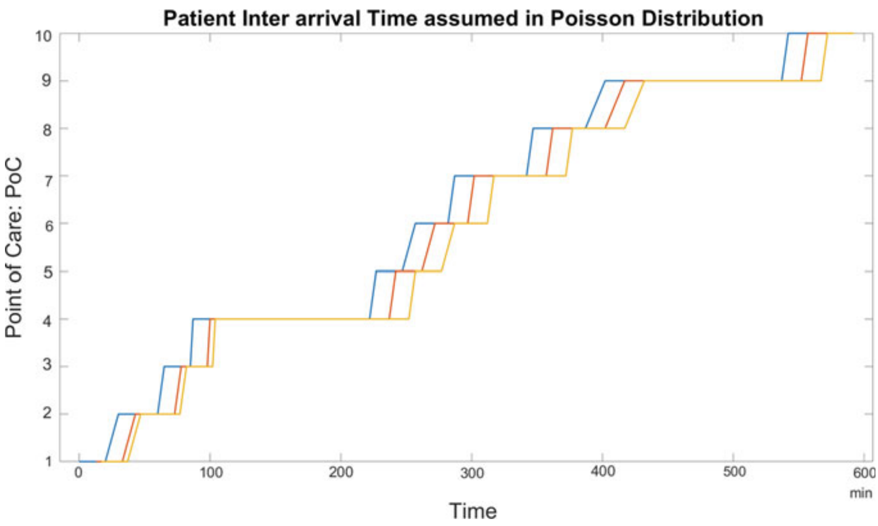


Fig. 8 The plot of the first three Patient's travel along the Point of Cares

Table 6 Increase in the improvement of process time by doubling the capacity only at PoC 4

Patient No	Arrival time (minutes)	Process end time (minutes) till PoC4, when only one bed is operated in each PoC	Process end time (minutes) till PoC4, when two beds are operated only in PoC 4	Percentage improvement
P1	0	132	132	NA
P2	13	134	132	2%
P3	17	145	138	5%
P4	24	153	141	9%
P5	36	156	139	12%
P6	47	160	138	16%
P7	64	158	132	20%
P8	68	169	138	22%
P9	76	176	140	26%
P10	93	174	133	31%

factor as per the standards of the hospital. For this purpose, the time slot for consideration can be three or four hours for scheduling the medical staff. In this manner, the dual resource scheduling based on patient arrival patterns can be achieved for that particular time interval.

6 Conclusions

In this paper, we have formulated a Mixed-Integer Linear Programming model for activation of the beds at each PoC. The optimization model's performance is compared with the practical approach followed by a trauma center in India. According to the results, the capacity enhancement at Critical PoCs can improve the hospital's performance in delivering timely care to the patients by optimizing the medical staff for that particular time interval. This helps in utilizing the scarce and valuable medical staff resources optimally while delivering quality and timely care to the Patients.

Further Research

The optimization can be included from the time of the accident, travel by ambulance, and completion of the operation theatre. The optimization can also consider the cost of enhancing capacity at a particular PoC in the workforce and additional equipment. The optimization can also be improved to evaluate the improvement in transit times and process times.

References

1. Gale, J., Coburn, A., Pearson, K., Croll, Z., Shaler, G.: Developing program performance measures for rural emergency medical services. J. Prehospital Emerg. Care, pp. 157–165 (2017)
2. Edwin, J.P.R.: Improving emergency department patient flow. Clin. Exp. Emerg. Med. **3**(2), 63–68 (2016)
3. Paul Richard, E.J.: Improving emergency department patient flow. Clin. Exp. Emergency Med. **3**(2), 63 (2016)
4. Fee, C., Weber, E.J., Maak, C.A., Bacchetti, P.: Effect of emergency department crowding on time to antibiotics in patients admitted with community-acquired pneumonia. Ann. Emerg. Med. **50**(5), 208 (2007)
5. Singer, A.J., Thode Jr, H.C., Viccellio, P., Pines, J.M., Schull M.J.: The association between length of emergency department boarding and mortality. Ann. Emerg. Med **44**(6), 577–85 (2004)
6. Diercks, D.B., Roe, M.T., Chen, A.Y., Peacock, W.F., Kirk, J.D., Pollack Jr, C.V., Gibler, W.B., Smith Jr, S.C., Ohman, M., Peterson, E.D.: Prolonged emergency department stays of non-ST-segment-elevation myocardial infarction patients are associated with worse adherence to the American College of Cardiology/American Heart Association guidelines for management and increased adverse events. Ann. Emerg. Med. **50**(5), 489–96 (2007)
7. Wiler, J.L., Griffey, R.T., Olsen, T.: Review of modeling approaches for emergency department patient flow and crowding research. Acad. Emerg. Med. **18**(12), 1371–1379 (2011)
8. Bernstein, S.L., Verghese, V., Leung, W., Lunney, A.T., Perez, I.: Development and validation of a new index to measure emergency department crowding. Acad. Emerg. Med. **10**(9), 938–942 (2003)
9. Stephen, L.T., Epstein, K.: Development of an emergency department work score to predict ambulance diversion. Acad. Emerg. Med. **13**(4), 421–426 (2006)
10. Nathan Hoot, D.A.: An early warning system for overcrowding in the emergency department. In: AMIA Annual Symposium Proceedings of American Medical Informatics Association. vol. 2006, p. 339 (2006)
11. Timothy, H.G.G., Reeder, J.: When the safety net is unsafe: real-time assessment. Acad. Emerg. Med. **8**(11), 1070–1074 (2008)
12. Weiss, S.J., Ernst, A.A., Derlet, R., King, R., Bair, A., Nick, T.G.: Estimating the degree of emergency department overcrowding in academic medical centers: results of the national ED Overcrowding Study (NEDOCS). Acad. Emerg. Med. **11**(1), 38–50 (2004)
13. Weiss, S.J., Ernst, A.A., Derlet, R., King, R., Bair, A., Nick, T.G.: Relationship between the National ED Overcrowding Scale and the number of patients who leave without being seen in an academic ED. Am. J. Emerg. Med. **23**(3), 288–294 (2005)
14. Batal, H., Tench, J., McMillan, S., Adams, J., Mehler, P.S.: Predicting patient visits to an urgent care clinic using calendar variables. Acad. Emerg. Med. **8**(1), 48–53 (2001)
15. Milner, P.C.: Forecasting the demand on accident and emergency departments in health districts in the Trent region. Stat. Med. 7(10), 1061–1072 (1988)
16. Milner, P.C.: Ten-year follow-up of ARIMA forecasts of attendances at accident and emergency departments in the Trent region. Stat. Med. **16**(18), 2117–2125 (1997)
17. Champion, R., Kinsman, L.D., Lee, G.A., Masman, K.A., May, E.A., Mills, T.M., Taylor, M.D.: Forecasting emergency department presentations. Aust. Health Rev. **31**(1), 83–90 (2007)
18. Sun, Y., Heng, B.H., Seow, Y.T., Seow, E.: Forecasting daily attendances at an emergency department to aid resource planning. BMC Emerg. Med. (2009)
19. A multi-class queueing network analysis methodology for improving hospital emergency department performance. Comput. Oper. Res. **36**(5), 1497–1512 (2009)
20. Mayhew, L., Smith, D.: Using queuing theory to analyse the government's 4-H completion time target in accident and emergency departments. Health Care Manage. Sci. **11**(1), 11–21 (2008)

21. Chakraborthy, N., Pandian, G.R., Dhanawade, V.S., Bhanu T.K., Priya, K., Abhilash, K.P.: Profile of trauma patients in the emergency department of a tertiary care hospital in South India. J. Family Med. Primary Care **5**(3), 558–563 (2016)
22. "KGMU Website," 2019. Online. Available: http://www.kgmu.org/hospital_trauma1.php
23. Pratorious, T., Chaudhuri, A., Saddikuti, V., Hasle, P., Singh, A.: Achieving better integration in Trauma care delivery in India: insights from a patient survey. J. Health Manage. **20**(3), 1–21 (2018)
24. Oestern, H.-J., Garg, B., Kotwal, P.: Trauma care in India and Germany. Clin. Orthop. Relat. Res. **471**, 2869–2877 (2013)
25. L. J. G. J. J. M. P. M. Mock, C.: Guidelines for essential Trauma care. Geneva. World Health Organization, pp. 4–5 (2004)

Real-Time Scheduling of Bed Resource Allocation to Improve Emergency Overcrowding

Shao-Jen Weng, Chih-Hao Chen, Yao-Te Tsai, and Shih-Chia Liu

Abstract Previous studies usually investigated improve the overcrowding in the emergency department (ED) as the research goal. In the case of limited medical resources, the congestion in the ED was due to the capacity of the hospital, and the imbalance of resources in the hospital is one of the bottlenecks. For this problem, we simulated the use of bed resources in different departments through discrete events, and dispatched the allocation of bed resources in real time. Finally, we compared three bed resource scheduling plans. This study shows that the best plan can improve the proportion of admitted patients held >48 h from the original 5.08 to 1.83%; The proportion of admitted patients held >24 h, stays dropped from 19.49 to 16.21%, and the variance of daily ED occupancy was reduced by 7%.

Keywords Overcrowding · Hospital operations · Emergency department · Simulation

1 Introduction

The ED is facing unprecedented pressure. If the ED is not invested or improved according to the increase of patients in the future, the waiting time of emergency patients may increase [1]. Congestion in the ED occurs because the number of inpatients exceeds the number of beds in the medical institution. It is not a specific problem in the emergency department, because the congestion in the ED is caused by the congestion of the medical institution [2], and the high bed utilization rate

S.-J. Weng · C.-H. Chen (✉) · S.-C. Liu
Department of Industrial Engineering and Enterprise Information, Tunghai University, Taichung, Taiwan
e-mail: p8311011@gmail.com

Y.-T. Tsai
Department of International Business, Feng Chia University, Taichung, Taiwan

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_81

may affect the waiting time of emergency patients and the organizational relationship between the emergency room and the ward, and emergency patients wait for a long time. It seems to be related to the poor performance of the medical system or long-term care management [3].

In order to increase the turnover rate of beds, the improvement of the flow of people and the ability of the medical staff to move, the early discharge of patients or the rapid bed cleaning, etc., are all easy to implement methods. After sorting, there are four intervention measures, namely, smoothing elective admissions, early discharge, weekend discharge and full capacity action plan, these methods require communication, cooperation and compromise between departments, in order to improve the overall capacity of the hospital and reduce the number of boarded patients in the ED [4], the external system impact of the ED is indispensable. The simulation must expand its research scope and check the system impact outside the ED itself [5].

The current research trend of unbalanced medical resources is the methodology for patient referral between graded hospitals [6–8], but there are also problems with imbalances in medical resources within the hospital, such as the imbalance of bed resources between departments, which causes overcrowding in the ED [9]. There is a close relationship between multiple departments of the hospital and the overcrowding of the ED, but most studies discuss the relationship between a single department and the ED [10–12], or the scheduling of beds based on patient types [13–15]. In addition, a specific indicator may have a negative impact on another indicator. Few studies are for the allocation of multi-sectoral resources for multiple indicators, but the combination of multi-objective methods or different improvement methods can improve the hospital more comprehensively [16].

When the hospital is overloaded, it will cause the ED to be overcrowded. Most hospitals will formulate surge agreements to deal with the measures when the hospital is overloaded. There are usually two types of surge agreements. One is the emergency response agreement for major disasters; One is the protocol of daily surge, including increasing emergency beds or allowing non-acute patients to be discharged, etc., to improve the daily overcrowding situation [17, 18], this study belongs to the latter.

Emergency response capability has been identified as a key indicator of hospital performance in the United States, and coordinating patient discharge and smoothing workload is usually an effective strategy to improve emergency response [19]. Hospital administrators can use bed scheduling measures to smooth the number of people in ED, improve surge capacity, and maintain ED quality indicators [9], there is no research to improve ED quality indicators through bed scheduling measures. Therefore, this study simulates the real-time scheduling of multi-departmental beds to alleviate the demand for emergency beds, put forward the method of real-time scheduling of beds and three strategies for multi-departmental scheduling of beds, How to improve ED quality indicators and solve ED overcrowding.

2 Materials and Methods

2.1 Model Development

This study is based on the boarding time data of emergency patients at Taichung Veterans General Hospital (TCVGH) from January to September 2018 and the types of patients who mainly use beds provided by the hospital, and construct models for emergency patients, elective surgery patients, invasive examination patients, medical examination patients, and patients transferred from the intensive care unit (ICU) to simulate the use of beds by different types of patients. When boarding patients are stranded in the ED, they need to have vacant beds before they can be hospitalized.

The total number of beds is set to 900, and there are 80 patients in the initial ED. It is assumed that distribution of the number of patients in different departments who need to be admitted to the hospital, the distribution of length of stay and the arrival time distribution of patients each day in Table 1, patients outside the emergency department have priority in using beds, because different departments have their own beds that can be used first. In addition, regardless of the need for additional beds in departments other than the emergency department or the difference in patient arrival time, they will be admitted to the hospital at 12 noon.

In this research, we use SIMUL8, a set of discrete event simulation software tools, to develop a real-time scheduling bed resource simulation model, and to track five indicators, namely the proportion of ED patients held >24 h, the proportion of ED patients held >48 h, the proportion of admitted patients held >24 h, the proportion of admitted patients held >48 h and the variance of daily ED occupancy, assess whether to improve ED overcrowding. In the past, data-driven discrete event simulation was used to help hospitals decide on overcrowded improvement plans [18]. Simulations are also applied to patient queuing, resource constraints, and multi-unit hospital models to evaluate hospital performance in different scenarios [20], and there are

Table 1 Fitting statistical distributions for different departments in the model

Department	Distribution of the number of patients (number of people)	Distribution of length of stay (min)	Arrival time distribution of patients (min)
Elective surgery	Uniform (120,140)	Exponential (4320)	Fix (1440)
Invasive examination	Uniform (1550)	Exponential (4320)	Fix (1440)
Medical examination	Uniform (4060)	Exponential (1440)	Fix (1440)
Transfers from the intensive care unit (ICU)	Uniform (1015)	Exponential (10,080)	Fix (1440)

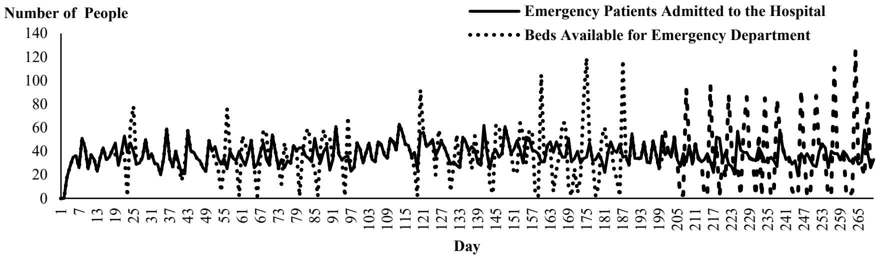


Fig. 1 Comparison of the number of boarded patients each day and the number of available hospital beds in the ED

many related literatures on optimizing bed capacity and allocation, reducing emergency waiting time and maximizing operating room utilization and other related resource allocation simulation models [21].

2.2 ED Situation

The average number of boarding patients per day is about 37, and the peak number is 63. The average number of beds provided ED is 37 beds per day, and the peak number of beds provided is 125 beds. The average daily difference between its hospital beds and emergency patients waiting to be hospitalized is 12.91, and the maximum difference is 90. It is often impossible to provide emergency beds to emergency patients immediately (see Fig. 1), causing patients to stay in the ED. The quality indicators that Taiwan hospitals pay attention to are the proportion of ED patients held >24 h, the proportion of ED patients held >48 h, the proportion of admitted patients held >24 h, and the proportion of admitted patients held >48 h. The current situation in order of 5.46, 1.42, 19.49 and 5.08%, if the number of available beds for emergency patients can be immediately supplemented, it can effectively reduce the proportion of ED patients held.

2.3 Analysis of the Strategy

In this study, based on the maximum number of stranded emergency patients, we decided whether to schedule beds, and set the number of days to return the scheduled bed, the number of scheduled beds at one time, and the maximum scheduling bed in Table 2, by sequentially changing the decision variables, find out the better setting of the hospital's scheduling beds, and compare the strategies of different departments' bed adjustment ratios in Table 3. The priority of the indicators is from the proportion of ED patients held >48 h, the proportion of admitted patients held >48 h, the proportion of ED patients held >24 h, the proportion of admitted patients held >24 h

Table 2 Decision variables of beds scheduling mechanism

	The maximum number of stranded emergency patients (number of people)	The number of days to return the scheduled bed (day)	The number of scheduled beds at one time (number of bed)	The maximum scheduling bed (number of bed)
Decision variables	90 100 110 120	1 2 3 4 5 6	20 30 40 50	20 30 40 50 60

Table 3 The strategies of different departments' bed adjustment ratios

Scenario	The ratio of scheduled beds in elective surgery (%)	The ratio of scheduled beds in invasive examination (%)	The ratio of scheduled beds in medical examination (%)	The ratio of scheduled beds in transfers from the intensive care unit (ICU) (%)
S1	100	0	0	0
S2	25	25	25	25
S3	58	14	22	6

and the variance of daily ED occupancy is compared by the importance of the index from high to low.

3 Results

Tables 4, 5, 6, and 7 show the simulation of different variables based on the maximum number of stranded emergency patients, the number of days to return the scheduled bed, the number of scheduled beds at one time, and the maximum scheduling bed. The average emergency quality index and the impact of the standard deviation (Std) of the number of people staying in the ED each weekly. The results show that setting the maximum number of stranded emergency patients at 100, the dispatched beds will be returned after 2 days, the dispatch of 40 beds at a time and the maximum scheduling bed at 50 beds, all indicators have better performance, compared with the original emergency quality indicators. The priority of the indicators is from the proportion of ED patients held >48 h, the proportion of admitted patients held >48 h, the proportion of ED patients held >24 h, the proportion of admitted patients held

>24 h and the variance of daily ED occupancy, the improvement ratio of each index was 58, 58, 13, 13 and 16%.

Tables 8 and 9 present the three multi-departmental strategies proposed based on the above-mentioned optimal scheduling bed variable settings and simulations. The results show that S2's the proportion of patients held are better than other strategies, although the variance of daily ED occupancy was reduced by 7%, which is lower than S1's 16% and S3's 13%, but judging from the comparison of hospital priority indicators, S2 is obviously the best among all emergency quality indicators. We

Table 4 The result of different variables based on the maximum number of stranded emergency patients

Maximum number of stranded emergency patients (number of people)	ED patients held >48 h	Admitted ED patients held >48 h	ED patients held >24 h	Admitted ED patients held >24 h	Standard deviation
90	0.96	3.41	4.89	17.44	5.01
100	0.86	3.07	4.84	17.27	5.10
110	0.84	2.99	4.93	17.58	5.36
120	0.87	3.12	4.99	17.81	5.28

Table 5 The result of different variables based on the number of days to return the scheduled bed

The number of days to return the scheduled bed (day)	ED patients held >48 h	Admitted ED patients held >48 h	ED patients held >24 h	Admitted ED patients held >24 h	Standard deviation
1	0.86	3.07	4.84	17.27	5.10
2	0.76	2.71	4.83	17.23	4.93
3	0.80	2.85	4.88	17.41	5.07
4	0.83	2.95	4.83	17.22	5.34
5	0.85	3.05	4.86	17.32	5.60
6	0.91	3.24	5.00	17.83	5.54

Table 6 The result of different variables based on the number of scheduled beds at one time

Number of scheduled beds at one time (number of bed)	ED patients held >48 h	Admitted ED patients held >48 h	ED patients held >24 h	Admitted ED patients held >24 h	Standard deviation
20	0.76	2.71	4.83	17.23	4.93
30	0.73	2.60	4.77	17.03	4.84
40	0.71	2.55	4.79	17.09	4.80
50	0.71	2.55	4.79	17.09	4.80

Table 7 The result of different variables based on the maximum scheduling bed

Maximum scheduling bed (number of bed)	ED patients held >48 h	Admitted ED patients held >48 h	ED patients held >24 h	Admitted ED patients held >24 h	Standard deviation
20	0.87	3.10	4.90	17.48	5.00
30	0.71	2.55	4.79	17.09	4.80
40	0.67	2.40	4.91	17.51	4.65
50	0.60	2.13	4.75	16.96	4.61
60	0.59	2.12	4.86	17.34	4.32

Table 8 Comparison of the proportion of patients held indicators for three multi-departmental bed Scheduling Strategies

Strategy	ED patients held >48 h	Admitted ED patients held >48 h	ED patients held >24 h	Admitted ED patients held >24 h
Original	1.42	5.08	5.46	19.49
S1	0.60	2.13	4.75	16.96
S2	0.51	1.83	4.54	16.21
S3	0.55	1.97	4.65	16.59

Table 9 Comparison of the variance of daily ED occupancy indicators for three multi-departmental bed scheduling strategies

Strategy	Mean	Standard deviation	Improve efficiency (%)
Original	91.09	5.51	baseline
S1	87.12	4.61	16
S2	86.21	5.10	7
S3	86.66	4.79	13

found that the three strategies of multi-departmental bed scheduling are all effective, but when the proportion of patients held drops to a certain level, it will be inversely proportional to the variance of daily ED occupancy, Perhaps because the number of patients arriving in the emergency department has changed greatly, in order to maintain the smoothness of the number of patients staying in the emergency department, some patients who will stay in the emergency department for more than 24 and 48 h have been sacrificed.

4 Discussion

Studies have shown that the emergency department quality indicators of the strategy of scheduling the same number of beds in different departments are better than other strategies. Although the variance of daily ED occupancy was reduced by 7% of all strategies, we found that the three strategies of multi-departmental bed scheduling are all effective, but when the proportion of patients held drops to a certain level, and the difference between different departments in the simulation is only the difference between the length of hospitalization and the number of beds required per day. It also reveals that when the hospital beds are expected to meet the needs of patients, the occupancy time of hospital beds is very important, because the multi-departmental bed allocation strategy sets the proportion of the number of scheduled beds. Therefore, unless the number of beds required per day is less than the number of scheduled beds, the number of beds is not sufficient for scheduling, but the settings of the three strategies are all in Within the range of the number of beds required per day, if the estimated bed occupancy time or some patients with a known hospitalization period are scheduled, the beds will be better used and the emergency quality indicators will be improved. We conducted research through simulation analysis, and we found that if the decision variables of real-time scheduling of beds are optimized through data simulation, the effect of the strategy that can assist in scheduling of beds is more significant.

5 Conclusions

The imbalance of resources in the hospital is one of the bottlenecks of the hospital's capacity. The contribution of this study is that we propose three strategies for real-time scheduling of multi-departmental beds, which can simultaneously reduce the variance of daily ED occupancy and the proportion of patients held, and reduce the problem of ED overcrowding.

References

1. Turner, A.J., Anselmi, L., Lau, Y.S., Sutton, M.: The effects of unexpected changes in demand on the performance of emergency departments. Health Econ. **29**(12), 1744–1763 (2020)
2. Salway, R., Valenzuela, R., Shoenberger, J., Mallon, W., Viccellio, A.: Emergency department (ED) overcrowding: evidence-based answers to frequently asked questions. Rev. Méd. Clín. Las Condes **28**(2), 213–219 (2017)
3. Vainieri, M., Panero, C., Coletta, L.: Waiting times in emergency departments: a resource allocation or an efficiency issue? BMC Health Serv. Res. **20**(1), 1–10 (2020)
4. McKenna, P., Heslin, S.M., Viccellio, P., Mallon, W.K., Hernandez, C., Morley, E.J.: Emergency department and hospital crowding: causes, consequences, and cures. Clin. Exp. Emerg. Med. **6**(3), 189 (2019)
5. Salmon, A., Rachuba, S., Briscoe, S., Pitt, M.: A structured literature review of simulation modelling applied to emergency departments: current patterns and emerging trends. Oper. Res. Health Care **19**, 1–13 (2018)
6. Li, N., Kong, N., Li, Q., Jiang, Z.: Evaluation of reverse referral partnership in a tiered hospital system–a queuing-based approach. Int. J. Prod. Res. **55**(19), 5647–5663 (2017)

7. Li, N., Teng, D., Kong, N.: Threshold control policy optimization for real-time reverse referral decision of chinese comprehensive hospitals. IEEE Trans. Autom. Sci. Eng. **16**(1), 45–60 (2018)
8. Wang, J.-J., Li, Z.-P., Shi, J.J.: Hospital referral and capacity strategies in the two-tier healthcare systems. Omega **100**, 102229 (2021)
9. Tsai, J.C.-H., Weng, S.-J., Liu, S.-C., Tsai, Y.-T., Gotcher, D.F., Chen, C.-H., Chou, C.-A., Kim, S.-H.: Adjusting daily inpatient bed allocation to smooth emergency department occupancy variation. Healthcare **8**(2), 78 (2020)
10. Wong, H.J., Wu, R.C., Caesar, M., Abrams, H., Morra, D.: Smoothing inpatient discharges decreases emergency department congestion: a system dynamics simulation model. Emerg. Med. J. **27**(8), 593–598 (2010)
11. Gür, Ş., Eren, T.: Application of operational research techniques in operating room scheduling problems: literature overview. J. Healthc. Eng. (2018)
12. Vancroonenburg, W., De Causmaecker, P., Spieksma, F., Vanden Berghe, G.: Scheduling elective patient admissions considering room assignment and operating theatre capacity constraints. In: Proceedings of the 5th International Conference on Applied Operational Research, Lecture Notes in Management Science, pp. 153–158 (2013)
13. Howell, E., Bessman, E., Kravet, S., Kolodner, K., Marshall, R., Wright, S.: Active bed management by hospitalists and emergency department throughput. Ann. Intern. Med. **149**(11), 804–810 (2008)
14. Landa, P., Sonnessa, M., Tànfani, E., Testi, A.: A discrete event simulation model to support bed management. In: 2014 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (simultech), pp. 901–912. IEEE (2014)
15. Landa, P., Sonnessa, M., Tànfani, E., Testi, A.: Multiobjective bed management considering emergency and elective patient flows. Int. Trans. Oper. Res. **25**(1), 91–110 (2018)
16. Vanbrabant, L., Braekers, K., Ramaekers, K., Van Nieuwenhuyse, I.: Simulation of emergency department operations: a comprehensive review of KPIs and operational improvements. Comput. Ind. Eng. **131**, 356–381 (2019)
17. Jenkins, J.L., O'Connor, R.E., Cone, D.C.: Differentiating large-scale surge versus daily surge. Acad. Emerg. Med. **13**(11), 1169–1172 (2006)
18. Busby, C.R., Carter, M.W.: Data-driven generic discrete event simulation model of hospital patient flow considering surge. In: 2017 Winter Simulation Conference (WSC), pp. 3006–3017. IEEE (2017)
19. Mills, A.F., Helm, J.E., Wang, Y.: Surge capacity deployment in hospitals: effectiveness of response and mitigation strategies. Manuf. Serv. Oper. Manag. **23**(2), 367–387 (2021)
20. TariVerdi, M., Miller-Hooks, E., Kirsch, T., Levin, S.: A resource-constrained, multi-unit hospital model for operational strategies evaluation under routine and surge demand scenarios. IISE Trans. Healthc. Sys. Eng. **9**(2), 103–119 (2019)
21. Busby, C.R., Carter, M.W.: Benefits of a broader view: capturing the hospital-wide impact of surge policies with discrete-event simulation. In: International Conference on Human-Centred Software Engineering, pp. 41–51. Springer (2019)

Basic Medical Equipment and Human Resource Allocation Models in Philippine Barangay Health Centers Using Integer Linear Programming

Gabriel C. Bucu, Glenn Gerald Castañeda, Keziah Marella Cueto, Clara Franchesca Mendoza, and Angeline Nicole Regalado

Abstract The Philippines is one of the populous nations in the world that suffers from the scarcity of healthcare resources. The research aims to increase the percentage of population accommodated by healthcare resources in barangay health centers annually using Integer Linear Programming (ILP). It will be used to address the problem of the insufficiency of healthcare resources in the Philippines which can also be used on different local barangay health centers. The study will be comprised of two models—the Human Resource Model and Basic Medical Equipment Model. The researchers also conducted three different cases for each model to test the feasibility and optimality of the model. For the first case, the base model, it is found that there is an infeasible solution with the data used from 2018 for the Human Resource Model. For the second case, decreasing the budget of 17% given to the Department of Health for 2019. This case decreased the budget allocated for the Human Resource of healthcare resources which makes the model more infeasible. For the third case, the surplus in the budget constraint of the Basic Medical Equipment Model will be used as an incremental for the budget constraint in the Human Resource Model to point out the lack of budget as a source of infeasibility for the first two models of Human Resource.

Keywords Integer linear programming · Operations research · Healthcare resources · Barangay health center · Optimization

1 Introduction

The World Health Organization (WHO) defines a healthcare system as "well-functioning" if it provides impartial access to quality healthcare regardless of pay dimensions while protecting them from the financial consequences of poor health. Healthcare in the Philippines does not meet these set standards [1]. Healthcare in the

G. C. Bucu (✉) · G. G. Castañeda · K. M. Cueto · C. F. Mendoza · A. N. Regalado
Departmetnt of Industrial Engineering, University of Santo Tomas, Manila, Philippines
e-mail: gcbucu@ust.edu.ph

C.-Y. Huang et al. (eds.), *Intelligent and Transformative Production in Pandemic Times*, Lecture Notes in Production Engineering,
https://doi.org/10.1007/978-3-031-18641-7_82

Philippines suffers from a shortage of human medical resources, especially doctors. This makes the system run slower and less efficiently. In addition to this problem, public health facilities are more run-down. These facilities have fewer medical staff, inferior facilities, and a lack of medical supplies. Due to the limited resources in the country, proper allocation of these resources is necessary for a development to a more efficient healthcare system and to reach those who need efficient and effective healthcare service.

The Philippines reached one hundred million (100,000,000) mark of population landing twelfth (12th) of the most populous nation in the world [2]. Due to this, healthcare is compromised with only 23,927 health facilities servicing the entire population [3]. In 2016, 59.2% or 342,705 patients have died having no medical attention or was left unattended (Philippine Statistics Authority, 2016). This information is especially highlighted knowing that as of 2018, 21.6% of the Filipino Population is located below the poverty line leaving 2,160,000 having access to only 2587 healthcare facilities that charge little to none (Philippine Statistics Authority, 2018). Rural areas are where people are, which is why local government units prioritize catering problems to these areas but this cannot be resolved immediately by the government as the population increases rapidly each year [4]. Live birth in the Philippines was recorded to be at 1,731,289 or 17 births per 1000 population. These accounts have truly made it difficult to address the growing concerns on the lack of medical resources to attend to all of the Philippines' populations

Operations Research has long been used for a fair and efficient allocation of various resources in various fields, one of which is the field of medicine. Europe, a continent having countries like France, Italy, United Kingdom, and Netherlands top the world rankings with the most effective healthcare system (World Population Review 2019), have found it effective to utilize Operations Research. They have been making use of the Data Envelopment Analysis through the Linear Programming Method.

2 Materials and Methods

2.1 Research Design

The research adopted an action research design which means that a problem was comprehended and understood, therefore developing a model as a form of an intervention plan. In the research's case, the problem addressed was the high rate of casualties in local barangay brought about by the lack of medical attention due to inadequate medical resources in each local barangay.

The model developed was an Integer Linear Programming model which is to address the problem of inadequacy of medical resources (barangay health workers and basic medical equipment). The model can be applied in different local barangay health centers in the Philippines to generate an optimal number of essential medical resources that should be in the local barangay health centers according to House Bill

No. 1339 of 2016 of the Republic of the Philippines which will maximize the Human Resource Capacity and Procurement of each Barangay Health Centers given the Budget set by the Department of Health considering constraints from data gathered on the selected rural community considering factors such as population size and an average number of patients in the barangay health centers that gets accommodated.

Another method used was the case study research design. Three (3) case studies were done to analyze possible situations as to why the model produces such solutions through sensitivity analysis.

2.2 Data Measures

The Two (2) Integer Linear Programming (ILP) models were formulated considering the following conditions Provided the population of CHO 1 of the City of Dasmariñas, the plan of maximizing the Human Resources Capacity and the procurement of Basic Medical Equipment given the budget is to be achieved through the design of the ILP model keeping in mind the effect of acquiring adequate numbers of medical resources (barangay health workers and basic medical equipment).

Through the consideration of the following factors, the Integer Linear Programming was formulated:

1. The population of City Health Office 1
2. The recommended ratio of nurses to a number of people by the House Bill 1339 of 2016
3. The recommended ratio of midwives to a number of people by the House Bill 1339 of 2016
4. The recommended ratio of barangay health worker to a number of people by the House Bill 1339 of 2016
5. The Basic Medical Equipment enumerated under the House Bill 1339 of 2016
    - Blood Pressure Apparatus.
    - Dressing Kit.
    - Weighing Scale.
    - Thermometer.
    - Glucometer.
    - Delivery Kit.

Decision Variables for the First Model

$\mathbf{X_1}$ = number of nurses needed to accommodate the total number of population covered by the barangay health center.
$\mathbf{X_2}$ = number of midwives needed to accommodate the total number of population covered by the barangay health center.
$\mathbf{X_3}$ = number of barangay health workers needed to accommodate the total number of population covered by the barangay health center.

The First Model is then written as follows:

$$\text{Max Z} = \binom{3}{i=1} x_i$$

Subject to:

$$X_1 \geq \frac{A}{5000}$$

$$X_2 \geq \frac{A}{2500}$$

$$X_3 \geq \frac{A}{500}$$

$$X_1 \leq 90038$$

$$X_2 \leq 43044$$

$$X_3 \leq 7923324$$

$$360.000X_1 + 205.155X_2 + 121.620X_3 \leq B$$

The data in Tables 1 and 2 measures were used in translating the model to Lingo:

Decision Variables for the Second Model

$\mathbf{X_1}$ = number of blood pressure apparatus needed to accommodate the total number of population covered by the barangay health center
$\mathbf{X_2}$ = number of dressing kits needed to accommodate the total number of population covered by the barangay health center.

Table 1 Decision variables for model 1

Decision variables	
Xi	Number of health workers needed to accommodate the total number of population covered by the barangay health center

Table 2 Other notions used for objective function for model

Other notions used for objective function	
A	Population of city health office 1 of the city of dasmariñas
B	Budget allocated by the Philippine government for medical human resource for city health office 1 of the City of Dasmariñas

X_3 = number of weighing scales needed to accommodate the total number of population covered by the barangay health center.
X_4 = number of thermometers needed to accommodate the total number of population covered by the barangay health center.
X_5 = number of glucometers needed to accommodate the total number of population covered by the barangay health center.
X_6 = number of delivery kits needed to accommodate the total number of population covered by the barangay health center.

The Second Model is then written as follows:

$$\text{Max Z} = \left(^{6}_{i=1} x_i\right)$$

Subject to:

$$X_1 \geq 52$$

$$X_1 \geq \frac{1}{6}C$$

$$X_2 \geq C$$

$$X_3 \geq 52$$

$$X_3 \geq \frac{1}{5}(C)$$

$$X_4 \geq 52$$

$$X_4 \geq \frac{1}{5}(C)$$

$$X_5 \geq 52$$

$$X_5 \geq \frac{1}{2}(C)$$

$$X_6 \geq D$$

$$0.300X_4 + 0.08X_5 + 0.200X_6 + 0.06X_7 + 1.000X_8 + 0.100X_9 \leq E$$

The data in Table 3 and 4 measures were used in translating the model to Lingo:

Table 3 Decision variables for model 2

Decision variables	
Xi	Number of basic medical equipment needed to accommodate the total number of populations covered by the barangay health center

Table 4 Other notions used for objective function for model 2

Other notions used for objective function	
C	Population of accommodated patients of the city health Office 1 of the city of dasmariñas
D	Average number of births under city health office 1 of the city of dasmariñas
E	Budget allocated by the Philippine government for basic Medical equipment and general medical facilities for city health office 1 of the city of Dasmariñas

3 Results

Through the Integer Linear Programming Model, provided by the data gathered from the data given by the City Health Office (CHO) 1 on the 26 barangays they are administering, and barangay health center regulations stated on the House Bill no. 1339 of 2016, the researchers were able to identify the optimal number of healthcare resources (Health Workers and Basic Medical Equipment) to maximize the Human Resource Capacity and Procurement of each barangay health center with the budget allocated by the Department of Health. The software LINGO was used to test its optimality and feasibility of the model. The model's sensitivity was also examined by testing three (3) cases based on the evident observations and clear analysis upon acquiring initial results from the model done.

3.1 Case 1. Base Model—Human Resource Model

The first case was the original model derived from the data gathered including the House Bill 1339. The model was formulated based on the following constraints: the number of nurses, midwives, and barangay health workers obtained from the House Bill 1339, the total number of nurses, midwives and barangay health workers available in the Philippines, and budget allotted by the national government to the local government unit for healthcare from the General Appropriations Act to maximize the Human Resource Capacity and Procurement of the Basic Medical Equipment based on the budget allocated by the Department of Health for each barangay health center (Table 5).

The base model for the Human Resource was used in order to find the optimal number of human resources (nurses, midwives, and barangay health workers). The model was run and the results are as in Fig. 1.

Table 5 Case 1 Input for human resource model

```
!Objective Function;
Max = X1 + X2 + X3;
!Constraints;
!Barangay Health Worker;
!Nurses;
X1 >= 479044/5000;
!Midwife;
X2 >= 479044/2500;
!Health Worker;
X3 >= 479044/500;
!HR Supply;
X1 <= 90038;
X2 <= 43044;
X3 <= 7923324;
!Budget Constraint (in thousands);
360.000*X1 + 205.155*X2 + 121.620*X3 <= 11738.472;
!Non Negativity Constraints;
X1 >= 0;
X2 >= 0;
X3 >= 0;
```

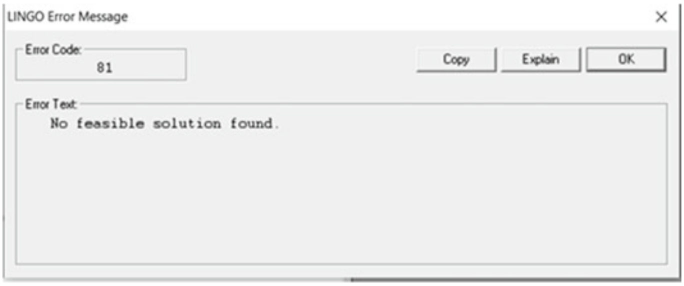


Fig. 1 Result of the case 1 model—human resource model

Data from 2018 was used in the model to obtain the result. Case 1 for Human Resource resulted in an infeasible solution. House Bill No. 1339 indicated a definite ratio of the medical personnel (number of nurses, wives, and barangay health workers) to the population which gives a definite number for these variables. The budget for the human resource allotted constrained the number of medical personnel that should be for the CHO 1 according to the standards in House Bill No. 1339. In analysis through arithmetic, the budget allotted for the human resource in the field of healthcare is not enough for the standards given by the house bill.

The base model for the Human Resource was used in order to find the optimal number of human resources (nurses, midwives, and barangay health workers). The model was run and the results are as in Table 6.

Table 6 Case 2 Input for human resource model

!Objective Function;
Max = X1 + X2 + X3;
!Constraints;
!Barangay Health Worker;
!Nurses;
X1 > = 486325/5000;
!Midwife;
X2 > = 486325/2500;
!Health Worker;
X3 > = 486325/500;
!HR Supply;
X1 < = 90038;
X2 < = 43044;
X3 < = 7923324;
!Budget Constraint (in thousands);
360.000*X1 + 205.155*X2 + 121.620*X3 < = 9742.932;
!Non Negativity Constraints;
X1 > = 0;
X2 > = 0;
X3 > = 0;

3.2 *Actual Budget Constraint—Human Resource Model*

The Model on Case 2 for Human Resource reflects the 17% decrease on the budget given to the Department of Health for 2019 by the Philippine Government allocating a higher percentage of 54% on the personnel services due to the implementation of the 2015 salary standardization law tranche 4 (Department of Health 2019).

The right-hand side of the model was changed due to an increase in population and a decrease in the budget for the year 2019.

Manipulation of the model only took place on the right-hand side assuming the population growth rate of 1.52% for the Philippines. The results of running the model for testing its feasibility are established as in Fig. 2.

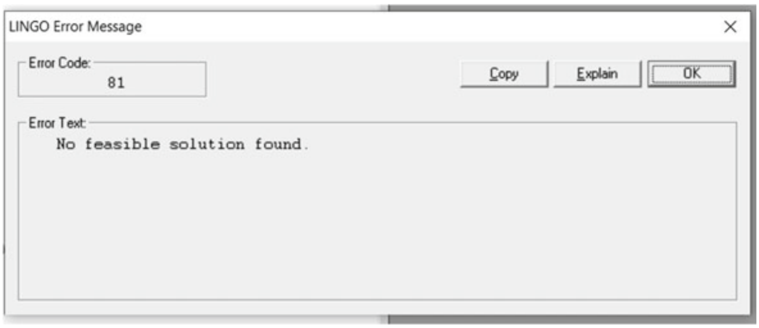


Fig. 2 Result of the case 2 model—human resource model

The outcome of the model revealed infeasible following a 17% decrease in the total budget even though there was a 54% increase in personnel service allocation. Employing the pattern of being infeasible on the base model, the decrease in budget while the population being serviced increases creates a great impact on the result.

3.3 Assuming 100 Percent Accommodation for CHO 1—Human Resource Model

Further model manipulation and analyzation were done to come up with a model that would exhibit a feasible solution, achieving the objective of maximizing the Human Resource Capacity of the Barangay Health Center in accordance with the budget allocated by the Department of Health. Capitalizing Case 3—Human Resources on satisfying the House Bill no. 1339's recommended healthcare resources (Health Workers, Basic Medical Equipment, General Medical Facilities) for barangay health centers to achieve 85% accommodation for the population of CHO 1, changes on the budget were made. The model for Case 3 is showcasing in Table 7.

Still using the projected 2019 population for City Health Office 1 the model on Case 3 for Human Resources increased further the budget by 1748%.

Applying the same method used on Case 2, the researchers retained the population and allocated 66% of the total budget allocated to Heath Service Budget to the Personnel Service.

Upon manipulation of the data, it was found that a surplus of ₱1,422,487.06 on the budget of equipment for 2019. Adding this up to the budget to be given, it was

Table 7 Case 3 input for human resource model

```
!Objective Function;
Max = X1 + X2 + X3;
!Constraints;
!Barangay Health Worker;
!Nurses;
X1 > = 486325/5000;
!Midwife;
X2 > = 486325/2500;
!Health Worker;
X3 > = 486325/500;
!HR Supply;
X1 < = 90038;
X2 < = 43044;
X3 < = 7923324;
!Budget Constraint (in thousands);
360.000*X1 + 205.155*X2 + 121.620*X3 < = 193217.896;
!Non Negativity Constraints;
X1 > = 0;
X2 > = 0;
X3 > = 0;
```

Table 8 Result of case 3 model—human resource model

Human resource	Amount to be hired
Nurses	98
Midwives	195
Barangay health worker	973
TOTAL	1265

computed that 66% of the budget is to be allocated to human resource. However, still appearing infeasible, the budget for Human Resource was raised to 1748% to get these results (Table 8).

3.4 Case 1. Base Model—Basic Medical Equipment Model

The Basic Medical Equipment base model was derived from the data included on the House Bill 1339. The model was formulated based on the following constraints: blood pressure apparatus, thermometer, glucometer, weighing scale, dressing kit and delivery kit obtained from the House Bill 1339, the useful lives of the equipment, the population of accommodation for City Health Office 1, and the budget allotted by the national government to the local government unit for healthcare from the General Appropriations Act to maximize Human Resource Capacity and Procurement of Basic Medical Equipment through the budget allocated to the barangay health centers by the Department of Health (Table 9).

The Basic Medical Equipment base model was used in order to find the optimal number of basic medical equipment (blood pressure apparatus, thermometer, glucometer, weighing scale, dressing kit and delivery kit). With the budget allocated to City Health Office 1, procurement for blood pressure apparatus is 1553, thermometer is 9318, glucometer is 1864, Weighing Scale is 25,576, Dressing Kit is 4659, and Delivery Kit is 4286, all of which are summarized in the Table 10.

3.5 Case 2. 2019 Actual Budget Constraint—Basic Medical Equipment Model

The Model on Case 2 reflects the 17% decrease on the budget given to the Department of Health for 2019 by the Philippine Government only allocating only 46% of the budget for the procurement of the Basic Medical Equipment (Table 11).

The right-hand side of the model were changed due to an increase in population and decrease on the budget for the year 2019.

The outcome of the model revealed the following after a 17% decrease in the total budget and allocating only 46% to the Basic Medical Equipment. The decrease

Table 9 Case 1 input for basic medical equipment model

```
!Objective Function;
Max = X1 + X2 + X3 + X4 + X5 + X6;
!Constraints;
!Basic Medical Equipment;
X1 >= 52;
X1 > = (1/6)*9318;
X2 > = 9318;
X3 > = 52;
X3 >= (1/5)*9318;
X4 > = 52;
X4 > = (1/5)*9318;
X5 > = 52;
X5 > = (1/2)*9318;
X6 > = 4286;
!Budget Constraints (in thousands);
0.300*X1 + 0.080*X2 + 0.200*X3 + 0.060*X4 + 1.000*X5
+ 0.100*X6 < = 8206.068;
!Non Negativity Constraints;
X1 > = 0;
X2 > = 0;
X3 > = 0;
X4 > = 0;
X5 > = 0;
X6 > = 0;
```

Table 10 Result of case 1 model—basic medical equipment model

Basic medical equipment to be procured yearly	
Budget	₱8,206,068
Blood pressure apparatus	1553 units
Thermometer	9318 units
Glucometer	1864 units
Weighing scale	25,572 units
Dressing kit	4659 units
Delivery kit	4286 units
TOTAL	47,254 units

in budget while the population being serviced increases creates great impact on the results summarized in the Table 12.

Table 11 Case 2 input for basic medical equipment model

```
!Objective Function;
Max = X1 + X2 + X3 + X4 + X5 + X6;
!Constraints;
!Basic Medical Equipment;
X1 > = 52;
X1 > = (1/6)*9318;
X2 > = 9318;
X3 > = 52;
X3 > = (1/5)*9318;
X4 > = 52;
X4 > = (1/5)*9318;
X5 > = 52;
X5 > = (1/2)*9318;
X6 > = 4286;
!Budget Constraints (in thousands);
0.300*X1 + 0.080*X2 + 0.200*X3 + 0.060*X4 + 1.000*X5
+ 0.100*X6 < = 6811.036;
!Non Negativity Constraints;
X1 > = 0;
X2 > = 0;
X3 > = 0;
X4 > = 0;
X5 > = 0;
X6 > = 0;
```

Table 12 Result of case 2 model–basic medical equipment model

Basic medical equipment to be procured yearly	
Budget	₱6,811,035.66
Blood pressure apparatus	1553 units
Thermometer	9318 units
Glucometer	1864 units
Weighing scale	2321 units
Dressing kit	4659 units
Delivery kit	4286 units
TOTAL	24,001 units

3.6 Case 3. Assuming 100% Accommodation for CHO 1—Basic Medical Equipment

Focusing Case 3 on the efficient allocation of the total budget for Barangay Healthcare, analysis on the surplus on each equipment without having to break the recommended number of Basic Medical Equipment by the House Bill no. 1339 was prioritized for better and fair allocation of budget. The model for Case 3 is showcasing (Table 13).

Table 13 Case 3 input for basic medical equipment model

```
!Objective Function;
Max = X1 + X2 + X3 + X4 + X5 + X6;
!Constraints;
!Basic Medical Equipment;
X1 > = 52;
X1 > = (1/6)*9318;
X2 > = 9318;
X3 > = 52;
X3 > = (1/5)*9318;
X4 > = 52;
X4 > = (1/5)*9318;
X5 > = 52;
X5 > = (1/2)*9318;
X6 > = 4286;
!Budget Constraints (in thousands);
0.300*X1 + 0.080*X2 + 0.200*X3 + 0.060*X4 + 1.000*X5
+ 0.100*X6 < = 6783.580;
!Non Negativity Constraints;
X1 > = 0;
X2 > = 0;
X3 > = 0;
X4 > = 0;
X5 > = 0;
X6 > = 0;
```

The researchers retained the population and the budget using the sensitivity analysis in reference to the 2018 budget and is then allocated to the Basic Medical Equipment. Projecting the possible budget increase and using the sensitivity analysis from the actual budget allocated to the Department of Health for 2018, a surplus of ₱1,422,487.06 was found and therefor was added to the Human Resource budget. The Procurement of Basic Medical Equipment are as follows (Table 14).

Table 14 Result of case 3 model–basic medical equipment model

Basic medical equipment to be procured yearly	
Budget	₱6,783,580.00
Blood pressure apparatus	1553 units
Thermometer	9318 units
Glucometer	1864 units
Weighing scale	1864 units
Dressing kit	4659 units
Delivery kit	4286 units
TOTAL	23,544 units

4 Discussion

4.1 Conclusion

The researchers formulated two (2) integer linear programming models for local government to use to determine the number of healthcare resources (nurses, midwives, barangay health workers, blood pressure apparatus, dressing kits, weighing scales, thermometer, glucometer, delivery kit) needed in their community to maximize the Human Resource Capacity and Procurement of the local barangay health centers. The first model is for human resource planning which takes into consider for constraints the supply of each human resource (nurses, midwives, and barangay health workers) available in the Philippines, the ratio mandated in the House Bill No. 1339, the salary given to them annually total budget allocated for Human Resource. The second one is the model for the procurement of basic medical equipment such as blood pressure apparatus, thermometer, glucometer, dressing kits and delivery kits which on the other hand takes consider as a constraint each equipment life cycle before repurchasing, minimum number of each equipment mandated by the House Bill No. 1339, and the price of each equipment correlated to the budget for equipment procurement.

Three (3) cases were analyzed for the model's sensitivity. First, the base model, which provided an infeasible solution for the human resource. Since there is a definite number of medical personnel needed, by doing basic analysis through arithmetic in the model, the budget for human resource does not suffice the given standards. The base model for the equipment procurement yielded a feasible solution but excessive than what's only needed. The second case was using the actual 2019 budget as budget constraint which still yielded an infeasible solution for the human resource. On the other hand, the model for the equipment is feasible yet there are still excess resources from the solution. The last model was based on the Case1: Human Resource Base Model and Medical Resources Base Model. This model led to still an infeasible solution for the human resource model and a feasible but with excess solution in the equipment procurement model. From the second solution of the model for equipment, a sensitivity analysis was done. From the sensitivity analysis, the considerable limit of the right-hand side for the budget was reduced from the budget allocation for procurement in the second model and was considered as surplus. This surplus for procurement was then added to the budget for the human resource but the model for the human resource is still infeasible. It was then found that a minimum increase of 1748% in human resource budget for it to be feasible. This proves that there is an imbalance in the distribution of the budget in human resource and procurement of equipment.

In conclusion with the different cases tested in the model, the budget highly influences the effectivity of the healthcare system in the Philippines; effective as defined by the Merriam-Webster Dictionary as producing a desired effect. The objective was to maximize the number of human resource capacity and procurement to accommodate the population of the barangay but due to the lack of budget, the model was not able to deliver an optimal solution. When the budget was manipulated, it then gained an optimal solution. This shows that the budget greatly affects the solution of the model which then translates that the budget allocation is not enough for the standards set by the government itself. Moreover, when further analyzed, the budget allocation for the human resource was the one that keeps the model infeasible which then again translates that the budget for human resource is short for the standards mandated by the government. However, when the desired objective was sought, the budget that came from the optimal solution was ideal but not realistic for the available resources of the Philippines. The results showed that the budget allocation for healthcare is not proportional to the needs of the population of the Philippines, hence, there is a clear image of inadequacy of resources in the country which reflects to the healthcare system.

4.2 Recommendations

Upon the analysis of the results from the study and correlating it to the current state of the country, the researchers recommend the Philippine Government:

- To reconsider to improve and prioritize the budget allocation on the healthcare system of the Philippines.
- To reevaluate the House Bill No. 1339 that mandated the standards for human resource and equipment.

Based on the results of the three (3) cases presented per model, the researchers recommend that the most effective model to be used is case 3 as it proposes a more feasible solution to the problem. It was the model that determined that there is a need to adjust the budget allocated for the human resource. Other models resulted a solution of either there is a surplus in equipment or shortage in budget.

References

1. Alam, K.T.: Performance of female volunteer community health workers in Dhaka urban slums. Soc. Sci. Med. **75**(3), 511–515 (2012)
2. An Act Granting Benefits and Incentives to Accredited Barangay Health Workers and for Other Purposes, Republic Act 7883 (Congress of the Philippines February 20, 1995)
3. An Act to Expand the Benefits of Barangay Officials to Include Barangay Health Workers, Senate Bill 261 (Eighteenth Congress 2019)
4. An Act to Improve and to Promote Quality Delivery of Health Services in Barangays, House Bill 1339 (Seventeenth Congress 2016)