ISSN 1854-6250

APEM journal

Advances in Production Engineering & Management

Volume 18 | Number 2 | June 2023

University of Maribor

Published by CPE
apem-journal.org

Advances in Production Engineering & Management

Identification Statement

ISSN 1854-6250 | Abbreviated key title: **Adv produc engineer manag** | Start year: 2006
ISSN 1855-6531 (on-line)

Published quarterly by **Chair of Production Engineering (CPE)**, **University of Maribor**
Smetanova ulica 17, SI – 2000 Maribor, Slovenia, European Union (EU)

Phone: 00386 2 2207522, Fax: 00386 2 2207990
Language of text: English

APEM homepage: apem-journal.org
University homepage: www.um.si

APEM Editorial

OPEN ACCESS

Subsidizer: The journal is subsidized by Slovenian Research Agency

Advances in Production Engineering & Management is indexed and abstracted in the **WEB OF SCIENCE** (maintained by **Clarivate Analytics**): **Science Citation Index Expanded**, **Journal Citation Reports** – Science Edition, **Current Contents** – Engineering, Computing and Technology • **Scopus** (maintained by **Elsevier**) • **Inspec** • **EBSCO**: Academic Search Alumni Edition, Academic Search Complete, Academic Search Elite, Academic Search Premier, Engineering Source, Sales & Marketing Source, TOC Premier • **ProQuest**: CSA Engineering Research Database – Cambridge Scientific Abstracts, Materials Business File, Materials Research Database, Mechanical & Transportation Engineering Abstracts, ProQuest SciTech Collection • **TEMA** (**DOMA**) • The journal is listed in **Ulrich's** Periodicals Directory and **Cabell's** Directory

University of Maribor
Chair of Production Engineering (CPE)

Advances in Production Engineering & Management

Volume 18 | Number 2 | June 2023 | pp 133–266

Contents

Scope and topics 136

Real-time scheduling for dynamic workshops with random new job insertions by using deep reinforcement learning 137
Sun, Z.Y.; Han, W.M.; Gao, L.L.

Genetic algorithm-based approach for makespan minimization in a flow shop with queue time limits and skip-ping jobs 152
Han, J.H.; Lee, J.Y.

Hybrid forecasting modelling of cost and time entities for planning and optimizing projects in the die-cast aluminium industry 163
Muñoz-Ibañez, C.; Chairez, I.; Jimenez-Martinez, M.; Molina, A.; Alfaro-Ponce, M.

When core sorting and quality grading is beneficial to remanufacturers: Insights from analytical models 175
Cao, X.H.; Shi, X.L.; Lan, H.J.; Huang, D.

Engineering-to-order manufacturing: A criticality analysis of key challenges and solutions based on literature review 187
Fortes, C.S.; Tenera, A.B.; Cunha, P.F.; Teixeira, J.P.

Ranking dominant losses in small and medium-sized enterprises (SMEs) in the context of the lean concept application 199
Kondić, V.; Maglić, L.; Runje, L.; Marić, D.

An improved discrete particle swarm optimization approach for a multi-objective optimization model of an urban logistics distribution network considering traffic congestion 211
Li, K.; Li, D.; Ma, H.Q.

Blockchain-based tripartite evolutionary game study of manufacturing capacity sharing 225
Wang, T.Y.; Zhang, H.

Enhancing automated defect detection through sequential clustering and classification: An industrial case study using the Sine-Cosine Algorithm, Possibilistic Fuzzy c-means, and Artificial Neural Network 237
Nguyen, T.P.Q.; Yang, C.L.; Le, M.D.; Nguyen, T.T.; Luu, M.T.

Simulation and Genetic Algorithm-based approach for multi-objective optimization of production planning: A case study in industry 250
Bojic, S.; Maslaric, M.; Mircetic, D.; Nikolicic, S.; Todorovic, V.

Calendar of events 263

Notes for contributors 265

Journal homepage: apem-journal.org

ISSN 1854-6250 (print)
ISSN 1855-6531 (on-line)

Published by CPE, University of Maribor.

Scope and topics

Advances in Production Engineering & Management (*APEM journal*) is an interdisciplinary refereed international academic journal published quarterly by the *Chair of Production Engineering* at the *University of Maribor*. The main goal of the *APEM journal* is to present original, high quality, theoretical and application-oriented research developments in all areas of production engineering and production management to a broad audience of academics and practitioners. In order to bridge the gap between theory and practice, applications based on advanced theory and case studies are particularly welcome. For theoretical papers, their originality and research contributions are the main factors in the evaluation process. General approaches, formalisms, algorithms or techniques should be illustrated with significant applications that demonstrate their applicability to real-world problems. Although the *APEM journal* main goal is to publish original research papers, review articles and professional papers are occasionally published.

Fields of interest include, but are not limited to:

Additive Manufacturing Processes
Advanced Production Technologies
Artificial Intelligence in Production
Assembly Systems
Automation
Big Data in Production
Block Chain in Manufacturing
Computer-Integrated Manufacturing
Cutting and Forming Processes
Decision Support Systems
Deep Learning in Manufacturing
Discrete Systems and Methodology
e-Manufacturing
Evolutionary Computation in Production
Fuzzy Systems
Human Factor Engineering, Ergonomics
Industrial Engineering
Industrial Processes
Industrial Robotics
Intelligent Manufacturing Systems
Joining Processes
Knowledge Management
Logistics in Production
Machine Learning in Production
Machine-to-Machine Economy
Machine Tools
Machining Systems
Manufacturing Systems
Materials Science, Multidisciplinary
Mechanical Engineering
Mechatronics
Metrology in Production
Modelling and Simulation
Numerical Techniques
Operations Research
Operations Planning, Scheduling and Control
Optimisation Techniques
Project Management
Quality Management
Risk and Uncertainty
Self-Organizing Systems
Smart Manufacturing
Statistical Methods
Supply Chain Management
Virtual Reality in Production

APEM journal

Advances in Production Engineering & Management | **ISSN 1854-6250**
Volume 18 | Number 2 | June 2023 | pp 137–151 | Journal home: apem-journal.org
https://doi.org/10.14743/apem2023.2.462 | Original scientific paper

Real-time scheduling for dynamic workshops with random new job insertions by using deep reinforcement learning

Sun, Z.Y.[a,b], **Han, W.M.**[a,*], **Gao, L.L.**[a]

[a]School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, P.R. China
[b]School of Software, Pingdingshan University, Pingdingshan, Henan, P.R. China

A B S T R A C T

Dynamic real-time workshop scheduling on job arrival is critical for effective production. This study proposed a dynamic shop scheduling method integrating deep reinforcement learning and convolutional neural network (CNN). In this method, the spatial pyramid pooling layer was added to the CNN to achieve effective dynamic scheduling. A five-channel, two-dimensional matrix that expressed the state characteristics of the production system was used to capture the state of the real-time production of the workshop. Adaptive scheduling was achieved by using a reward function that corresponds to the minimum total tardiness, and the common production dispatching rules were used as the action space. The experimental results revealed that the proposed algorithm achieved superior optimization capabilities with lower time cost than that of the genetic algorithm and could adaptively select appropriate dispatching rules based on the state features of the production system.

A R T I C L E   I N F O

Keywords:
Real-time scheduling;
Machine learning;
Deep reinforcement learning (DRL);
Spatial pyramid pooling layer;
Artificial neural networks (ANN);
Convolutional neural networks (CNN)

**Corresponding author:*
wlmh63@163.com
(Han, W.M.)

Article history:
Received 7 December 2022
Revised 6 June 2023
Accepted 25 June 2023

1. Introduction

With the rapid development of information technology, China is gradually entering intelligent industry 4.0 [1]. With changing market demand, multiple-product small-batch order-type production has become prevalent in the manufacturing industry. Therefore, achieving sustainable and efficient workshop production under complex production conditions, responding quickly to market changes, and satisfying the diverse needs of customers are critical.

Although job shop scheduling problems (JSSP) [2, 3] primarily address static scheduling issues, many real-time disruption factors, such as equipment failure, dynamic order arrival, emergency order insertion, in the production process are ignored [4, 5]. The insertion of a new order can drastically change the number and mode of processing tasks. Error accumulation renders existing production scheduling schemes ineffective, which results in the failure of the production planning system [6]. Therefore, dynamic real-time production scheduling on the insertion of new jobs is crucial for timely response to disruption events and ensuring production requirements are satisfied.

Current dynamic scheduling methods under order disturbance include heuristic algorithms [7-11] and dispatching rules [12]. Wang *et al*. [13] proposed an improved particle swarm optimization (PSO) algorithm to solve the dynamic job shop scheduling problem. Caldeira *et al*. [14] solved the flexible job shop scheduling problem on the arrival of a new job by using the improved backtracking search optimization algorithm that minimized makespan, energy consumption, and system stability. Ghaleb *et al*. [15] proposed three heuristic algorithms to address the real-time scheduling problem when new jobs are added and equipment fails. Tang *et al*. [16] considered minimum energy consumption and makespan as optimization objectives and proposed an improved PSO algorithm to solve the dynamic scheduling problem of flexible flow shops under new job arrival and equipment failure. In most heuristic algorithms, the dynamic scheduling problem is converted into a multi-stage static problem. Short-sightedness appears as the disturbance scale increases. The dispatching rules method can immediately respond to dynamic disturbance events and exhibits short computing time and high solution efficiency.

Hundreds of dispatching rules have been proposed for shop scheduling [17, 18]. Zhang *et al*. [19] proposed a job shop dispatching rule selection system based on semantics to achieve adaptive selection of dispatching rules by scheduling objectives. To reduce the time of job completion and complexity of process design in the conventional dispatching rule design process, Zhang *et al*. [20] proposed an improved genetic programming algorithm that evolves effective dispatching rules automatically. Ferreira *et al*. [21] combined machine learning with the problem domain reasoning to generate effective dispatching rules. Although dispatching rules can respond to dynamic disturbance events in real time and exhibit short computing time and high solving efficiency, these methods are prone to local optimum and cannot be adjusted adaptively to respond to various production states.

Reinforcement learning (RL) [22] has been widely used in production scheduling because of its excellent optimization ability and high computational speed. The continuous interaction between agent and environment maximizes cumulative rewards [23]. Dispatching rules can be dynamically and flexibly selected based on the real-time production status, which is suitable for the dynamic production scheduling problems. Wang *et al*. [24] used Q-learning to train a single machine agent and realized the dynamic selection of the three basic dispatching rules with the minimum average tardiness as the optimization objective. Chen *et al*. [25] proposed a rule-driven method for generating high-quality composite dispatching rules for the multi-objective dynamic job shop scheduling problem by using the Q-learning algorithm. Qu *et al*. [26] proposed a Q-learning algorithm that solves the dynamic job shop scheduling problem under random job arrival and equipment failure by combining the neighborhood search algorithm with the Q-factor. Although the conventional reinforcement learning algorithm has achieved excellent results in solving dynamic production scheduling problems, the algorithm is limited to situations in which the dimension and scale of the system state space are small and discrete. In the deep reinforcement learning (DRL) [27] algorithm, the perception ability of deep learning is effectively combined with the decision-making ability of reinforcement learning. Thus, DRL can effectively performs complex decision-making in the high-dimensional state space. Zhu *et al*. [28] proposed a proximal policy optimization (PPO) algorithm to solve the flexible flow shop scheduling problem by minimizing makespan. The PPO algorithm outperformed the conventional heuristic algorithm in terms of the quality of the solution. Luo *et al*. [29] proposed a deep Q-network (DQN) algorithm to solve the real-time workshop scheduling problem with dynamic job arrival to minimize total tardiness and achieved excellent results in the randomly generated data experiment. Yang *et al*. [30] used the A2C algorithm to train an intelligent model for the permutation flow job shop scheduling problem. This model outperformed the conventional heuristic algorithm in terms of the solving time and solution quality. Li *et al*. [31] proposed a hybrid DQN (HDQN) algorithm to solve the dynamic flexible job shop scheduling problem under transportation resource constraints. In most studies, the production system state is expressed through numerical features, which requires special manual design.

The CNN is used for the feature extraction of system state features expressed by multi-channel images to effectively reduce manual design difficulty and exhibits excellent generality. However, because of the limitation of CNN feature extraction on the input size of training images, static scheduling is widely used. Liu *et al*. [32] designed three channels of two-dimensional data matrices as system state features for job shop scheduling problems and solved these problems using the AC algorithm to achieve excellent results on benchmark examples. Han *et al*. [33] combined the CNN and DRL to achieve dynamic job shop scheduling. Wang *et al*. [34] used the PPO algorithm to solve the job shop scheduling problem outperform GA. Subsequently, random fine-tuning was performed on some examples to test the generalization ability of the model.

This method is simple and produces excellent results by describing the state features of the production system using multi-channel images. However, because of the structural characteristics of the CNN, applying the method to dynamic variable data of various image sizes is difficult. The static model can only process data of the same size, and it exhibits limited generalization. Multi-channel image system feature representation is yet to be used for dynamic scheduling. Therefore, in this study, an SPP layer [35] was added to the last layer of the CNN so that the neural network can handle any size of the input image information and achieve dynamic scheduling under the state image expression mode of the production system.

The contributions of this paper are as follows: (1) This study is the first to use SPP with neural networks to solve the dynamic scheduling problem to allow the system model to handle the input state data of any scale; (2) The state feature expression mode of the production system was improved. The limitations of the conventional three-channel image design was overcome, and the system state feature was represented by five-channel data, considering both the global and local features of system processing state; (3) To assess the effect of the dispatching rules selected at each decision point on the overall scheduling objective, a novel reward function targeting total tardiness was developed; (4) Comprehensive parameter sensitivity experiments and algorithm result comparisons were performed. The effectiveness of the calculation speed and optimization ability of the proposed scheme was demonstrated. The method achieved excellent generalization results.

2. Overall scheduling framework

A dueling double DQN (DDDQN) algorithm framework was proposed to achieve dynamic job shop real-time scheduling with constant random new order insertion. The neural network of the DDDQN algorithm consists of a Q-network and a target network. Each network exhibits the same structure and is composed of the convolution layer and a full connection layer. To address the problem of the dynamic image size change caused by the dynamic order arrival, an SPP layer was added between the CNN convolution layer and full connection layer to ensure consistent output size of the CNN convolution layer. Thus, the scheduling problem was transformed into a multi-stage decision-making process by designing a state feature, action space, and reward function. The agent is trained through interaction with the environment, and the trained agent is applied to solve the online problem. The framework includes two parts, namely offline training and online application (Fig. 1).

The scheduling environment comprises equipment and order in the production system, which is used for the interaction with the agent and providing the current production system status information. The agent outputs the most appropriate dispatching rule and selects the highest priority operation for processing. The production system then enters the next state.

In the offline training phase, intermediate data generated in the learning process is stored in the replay memory, and a minibatch number of sample data are randomly sampled for training. The Q-network and the target network calculate the Q and target values of the system state, respectively. Q-network parameters are updated by the loss function calculated by the target value and Q value. The parameters of the target network are copied from the Q-network after a certain number of steps. The optimal action is selected according to the result of the Q value.

Although offline scheduling requires considerable time to train an agent, when an agent learns good policy, it can be widely used in online actual data scheduling to obtain optimal scheduling results rapidly. The execution process only requires the Q-network to calculate the optimal Q value without updating various network parameters, calculating the reward value, or storing sample data and other operations.

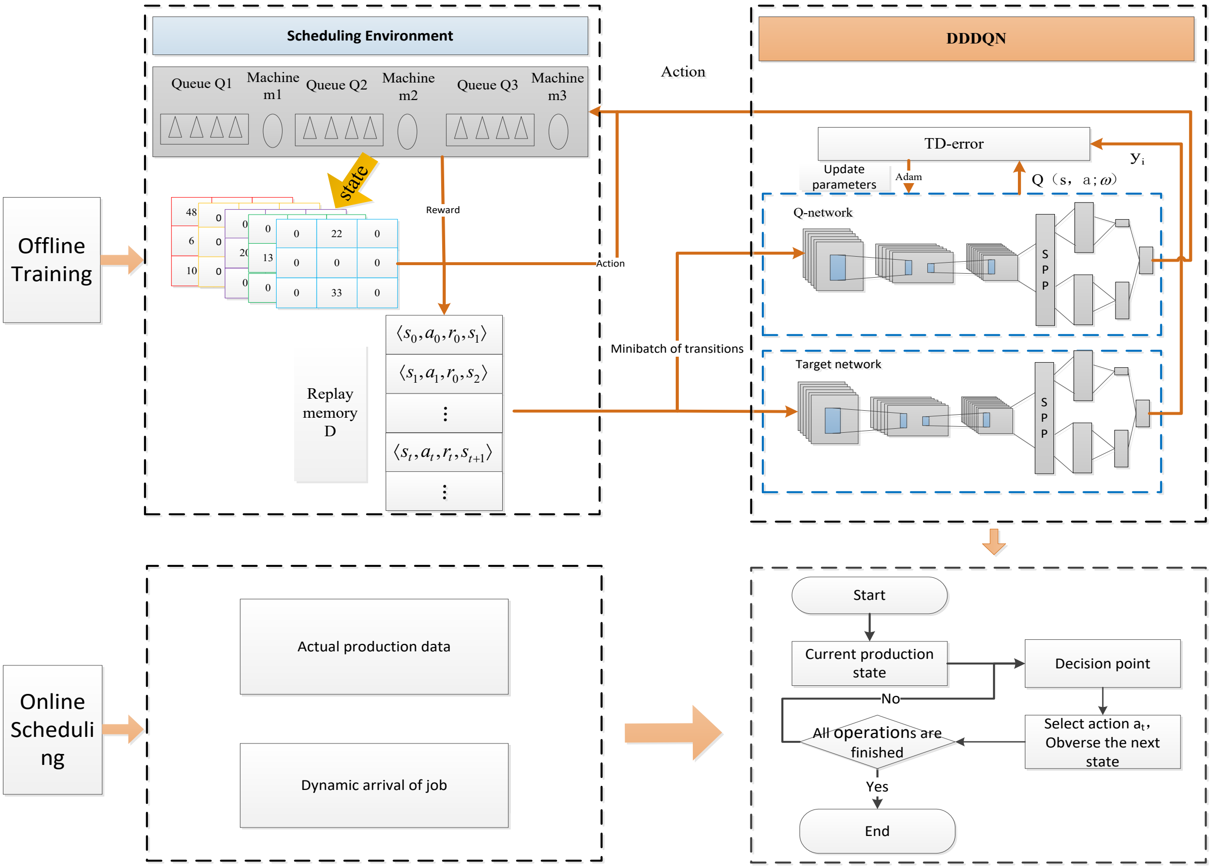


Fig. 1 Scheduling framework with the DDDQN

3. DRL for scheduling

3.1 Problem formulation

The JSPP with new order insertions can be described as follows: a processing system has N orders that are processed on M machines. Each order has n_j operations. The objective of production scheduling is to generate an optimal scheduling scheme based on the scheduling objectives and satisfy all constraints. However, when the new order arrives, the operations that were not started in the original scheduling scheme should be combined with operations in the new order for rescheduling. When creating the new scheduling scheme, factors such as the starting processing times and the number of the remaining operations of each order, should be considered. The scheduling problem should satisfy the following assumptions: (1) each order has a sequence constraint on the operations, that is, the next operation can only be processed after the previous operation is completed; (2) each operation of each order can only be processed by one machine; (3) each machine can only perform one operation at the same time; (4) when the new order arrives, the ongoing operation cannot be interrupted; (5) the processing time of each operation on the corresponding machine is known.

3.2 DQN principle

The DQN algorithm is used to solve the problem. The DQN is the most classical algorithms of DRL. Based on Q-learning, the deep neural network is used to represent value function f. The input of the neural network is the current state s, and the output is the state value function Q(s, a). In the DQN, empirical data are used to train the neural network, which is prone to instability and convergence difficulties. To solve these problems, replay memory and target network are used in the DQN. Experience replay stores the intermediate data in a fixed-size storage experience pool in the form of $\langle s_t, a_t, r_t, s_{t+1}\rangle$, which is generated in the learning process of Q-learning. The system randomly samples a certain number of small-batch samples from the replay memory for training. This random sampling not only breaks the correlation of training samples but also ensures an independent and homogeneous distribution of training samples. Target network reconstructs a network with the same structure as the original network. The original network is the Q-network, and the generated network is the target network. During training, only Q-network parameters are updated, and the parameters of the target network remain unchanged temporarily. After reaching a certain number of update steps C, the parameters of the Q-network are copied to the target network so that the value of the target network does not change in a certain update step to ensure system stability. The target value is calculated as follows:

$$Y^{DQN} \equiv r_{t+1} + \gamma \max_a \hat{Q}(s_{t+1}, a; \omega_t^-) \tag{1}$$

The neural network calculates TD errors through the target network and Q-network to update parameters.

In the standard DQN algorithm, the action with the largest Q value is selected, which results in an overestimation of the Q value. Therefore, the Double DQN (DDQN) algorithm is used to separate the action selected from the calculation of the Q value and two value functions are used. The target value of the DDQN is calculated as follows:

$$Y^{DoubleDQN} \equiv r_{t+1} + \gamma \hat{Q}(s_{t+1}, arg \max_a Q(s_{t+1}, a; \omega_t), \omega_t^-) \tag{2}$$

By using various value functions to decouple the target action and Q value, the DDQN algorithm mitigates the overestimation of Q values and achieves excellent stability.

In the DQN, the network outputs the Q value of action, but in practice, the Q value is associated with the action and state. Therefore, Dueling DQN improves the network structure of the DQN by adding state value function V(s, ω, a) related to the system state and the advantage function A(s, a, ω, β) related to the action before the network output layer and synthesizing these two functions to generate the action function in the final output layer. Thus, we have the following expression:

$$Q(s, a, \omega, \alpha, \beta) = V(s; \omega, \alpha) + (A(s, a, \omega, \beta) - \frac{1}{|\mathcal{A}|}\sum_{a_{t+1}} A(s, a_{t+1}, \omega, \beta)) \tag{3}$$

where ω is a parameter of the common part, α is a parameter of the value function, and β is a parameter of the advantage function.

In this study, dueling DQN and double DQN are used to solve the dynamic production scheduling problem.

3.3 CNN with the SPP layer

The insertion of new orders dynamically changes the size of multi-channel images expressed by state features. However, the full connection layer in the conventional CNN requires an input of fixed size. An SPP layer was added between the last convolutional layer and the first full connection layer in the CNN to divide the feature graph obtained after convolution into a fixed number of grids of various sizes. The grid is then pooled with mean values. Thus, the feature graph convolved with any size can be changed into the output of a fixed size so that the graph has the same dimension of feature vector with the following full connection layer. Thus, image convolution with any image size input can achieved as follows (Fig. 2).

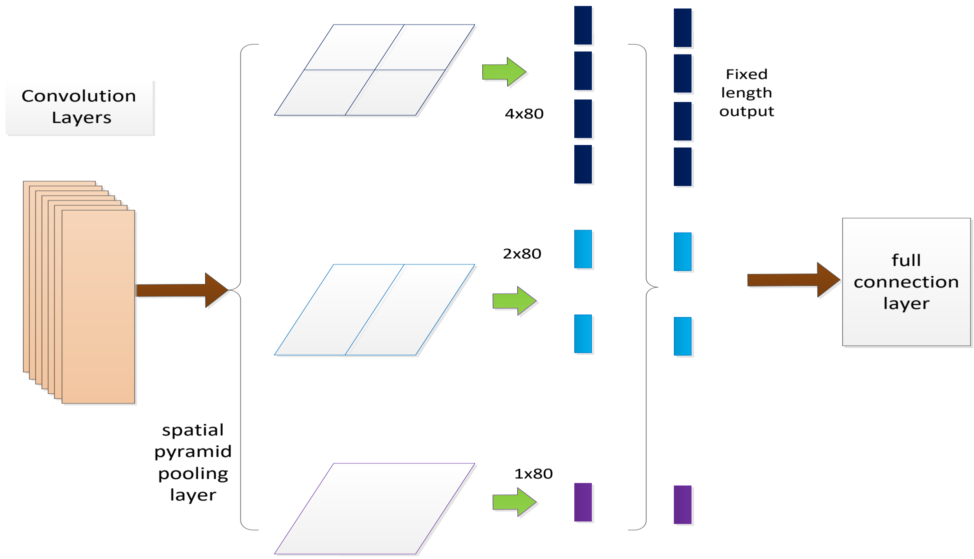


Fig. 2 SPP principle

3.4 Scheduling problem transformation

The problem transformation between scheduling and algorithm design is critical for applying DRL to JSPP and involves three aspects, namely state feature, action space, and reward function design.

State features expression

To improve the state changes of the production system, the following rules should be followed for describing the state features:

- State features should be able to reflect the features and changes of the production system, and both global and local state features should be considered.
- State features at each moment are represented by a universal feature set.
- State features should be represented numerically for easy calculation and standardization for uniform scaling of various features.

This study optimized and upgraded the production system state features based on literature [33]. The limitation of three channels in the conventional system state feature expression was overcome. Five channels were designed for characterization. The first channel is represented by the two-dimensional matrix of the order to be processed. The rows of the matrix represent the order, and the columns represent the operation. The initial value is the processing time of the corresponding operation in the order. The value of the corresponding position becomes zero on the completion of an operation. The second channel is represented by the two-dimensional matrix of the completed operations of the order. The initial value is zero. The value of the corresponding position is the processing time on completion of an operation. The third channel is the remaining processing time of the processing operation. The fourth channel is the processing time of each operation in the queue to be processed. The fifth channel is the waiting time of each operation in the waiting queue. The first and second channels express the global state feature, whereas the third, fourth, and fifth channels express the local state feature. All channel data are linearly normalized to the maximum value.

Definition of ACTIONS

The action selection involves selecting the most suitable operation waiting for processing, and the production scheduling rule can select an appropriate process at each scheduling decision point. In this article, 16 commonly used production dispatching rules are selected as the action space in DRL. The details are as follows: the select the job with the shortest processing time (SPT), select the job with the longest processing time (LPT), select the job with the longest remaining processing time (LWKR), select the job with the shortest remaining processing time (MWKR), select the job with the shortest processing time of subsequent operation (SSO), select the job with the longest processing time of subsequent operation (LSO), select the job with the shortest remaining processing time in addition to the current operation (SRM), select the job with the longest remaining processing time in addition to the current operation (LRM), select the job that arrives first (FIFO), select the job with the earliest due date, select the job with the minimum sum of processing time of the current and subsequent operation (SPT+SSO), select the job with the maximum sum of processing time of the current and subsequent operation (LPT+LSO), select the job with the minimum ratio of current processing time to the total working time (SPT/TWK), select the job with the maximum ratio of current processing time to the total working time (LPT/TWK), select the job with the minimum product of the current processing time and total working time (SPT *TWK), select the job with the maximum product of current processing time and total working time (LPT * TWK). The diversity of dispatching rules is increased so that the agent can adaptively select dispatching rules.

Reward FUNCTION

The reward function is key to DRL and directly affects the direction of learning and is closely related to optimization. The reward function should be designed as follows: (1) the reward function should accurately express the immediate reward of the current action. (2) Cumulative reward should be closely related to the scheduling objective. (3) Reward function should be universal and can be used for problems of various scales. Because the overall scheduling objective is to minimize total tardiness, the following reward function should be designed:

$$r_k = Tard_{k-1} - Tard_{k}, \qquad Tard_k = \sum_{j=1}^{n} \delta_j(\tau) \tag{4}$$

where $\delta_j(\tau)$ represents the tardiness of job J at the current system state, $Tard_k$ is the total tardiness of all the jobs in the current system. For the job without tardiness, the tardiness is zero. Here, r_k represents the reward received at the decision-making time t_{k-1} after executing the action, then the system arrives at decision-making time t_k. The derivation process of cumulative reward function R is as follows:

$$\begin{aligned} \mathrm{R} &= \sum_{k=1}^{K} r_k = \sum_{k=1}^{K} Tard_{k-1} - Tard_k \\ &= Tard_0 - Tard_1 + Tard_1 - Tard_2 \\ &\quad + \cdots + Tard_{K-1} - Tard_K \\ &= Tard_0 - Tard_K = -Tard_K \end{aligned} \tag{5}$$

The derivation process of the cumulative reward formula reveals that it is inversely proportional to the total tardiness, that is, the smaller the total tardiness is, the larger the cumulative reward value is, which is consistent with the scheduling objective.

Training based on DRL

The scheduling process is a semi-Markov decision process. When any machine completes an operation or a new order arrives as a decision time point, the agent adaptively selects appropriate dispatching rules, and the highest priority operation is selected for processing. Subsequently, the machine enters the next state after receiving rewards. The cycle continues until all operations are finished, that is, a scheduling scheme is obtained. The process is as follows:

Algorithm 1 DDDQN-based training method

```
Initialize replay memory D, minibatch k,  learning rate α,  target network parameters update every C steps.
Initialize Q-network with random weights ω
Initialize target network Q̂ with weights ω⁻ = ω
For episode = 1 : M do
Reset the system scheduling status to s_0 and clear schedule results.
while True : (t is the decision time point at which an operation is completed or a new order arrives, the
        Boolean variable done terminates the loop when all operations are complete)
```

7: Select action a_t based on ε-greedy strategy
8: Execute action a_t, calculate the immediate reward r_t, observe the next state s_{t+1}
9: Store transition$\langle s_t, a_{t,} r_t, s_{t+1}, done \rangle$ in D
10: Random sampling minibatch of transitions $\langle s_i, a_{i,} i, s_{i+1}, done \rangle$ from D
11: $y_i = \begin{cases} r_i & \text{if done} = \text{true} \\ r_i + \gamma \hat{Q}(s_{i+1}, argmax_a(s_{i+1}, a; \omega), \omega^-) & \text{otherwise} \end{cases}$
12: Compute TD-error$(y_i - Q(s_i, a; \omega))^2$to update the parameters of Q-network
13: Every C steps update the parameters of the target network $\hat{Q}$: $\omega^- = \omega$
14: end while
15: end for

The training process is divided into inner and outer loops, the outer loop represents the times of training. After M loops of cyclic training, the agent gradually reaches the ability to adaptively select the optimal dispatching rules at various decision moments. The inner loop represents a complete scheduling scheme generation process, starting from the first operation until all operations are finished, which is an episode, lines 5-14 describe the execution of the inner loop that starts from the initial state of s_0, at the decision moment t the agent select and execute the action a_t based on the ε-greedy strategy, the suitable operation is scheduled. The reward r_k and the next state s_{t+1} are observed, the transition $\langle s_t, a_{t,} r_t, s_{t+1}, done \rangle$ is stored. Minibatch transitions are randomly sampled from the replay memory for system training. A loss was calculated according to lines 11 and 12, and gradient descent was used to update the parameters of Q-network. The parameters of the target network were updated by the parameters of Q-network according to the contents of 13 lines after every C step.

4. Numerical simulation

In numerical simulation, multiple groups of data were used to train the DDDQN. The optimal training model is then saved, and the model is tested in the new test data of multiple groups of various production scenarios. The data generation method proposed in a previous study [29] is randomly generated, and the parameters are presented in Table 1.

According to the arrival time and number of new orders, the number of machines, and the due date of orders, 81 groups of data of various production scenarios were randomly generated for the test. Assuming that 30 orders exist at the beginning of each production scenario, the arrival time of new jobs follow Poisson distribution. Therefore, the time interval between two consecutive new jobs is subjected to exponential distribution. The due date tightness (DDT) represents the urgency of orders. The due date of order i can be calculated as $D_i = A_i + (\sum_j^{n_i} t_{ij}) \cdot DDT$. Here, A_i is the arrival time of the order, n_i is the number of operations in the order, t_{ij} is the processing time of the j operation of the order. The smaller the DDT is, the more urgent the order due date is.

Table 1 Parameter settings of various production scenarios

Parameter	Value
Number of machines	5,10,15
Number of initial jobs	30
Number of new job insertions	10,30,50
Processing time of each operation	Unif[0,50]
Due date tightness	1.0,1.5,2.0
Average value of exponential distribution between two successive new job arrivals	25,50,100

4.1 Network structure and system parameter setting

The CNN structure of the DDDQN consists of four convolution layers and two full connection layers. To solve the problem of variable image size, a SPP layer was added between the convolution layer and the full connection layer. From the first layer to the fourth layer for the convolution layers, the size of the convolution kernel was 6 × 6, 4 × 4, 3 × 3, 2 × 2, the step size was 2, 2, 2, 1, and the number of output channels was 20, 40, 60, and 80. Because each element in the state feature image represents an operation, the pooling layer in the CNN results in incomplete scheduling information. Therefore, the pooling layer was not used. The full connection layer consists of two branches of the full connection layer with 512 units. The two branches connect the state value and the advantage value. Finally, the state and advantage values were combined to obtain the final result output. The RELU activation function was used for every layer. The Adam optimizer was used to update the parameters.

Parameter setting considerably affected DDDQN performance. Five groups of data were randomly generated for the parameter sensitivity experiment under 10 new order insertions, 10 machines, 50 average time interval between new order arrival, and DDT tardiness coefficient of 1.5. The effects of the training batch, learning rate, replay memory buffer size, and target network parameter updating frequency on algorithm performance were verified. Fig. 3 displays the training effect under various parameter settings, the total number of training was set to 3000 episodes.

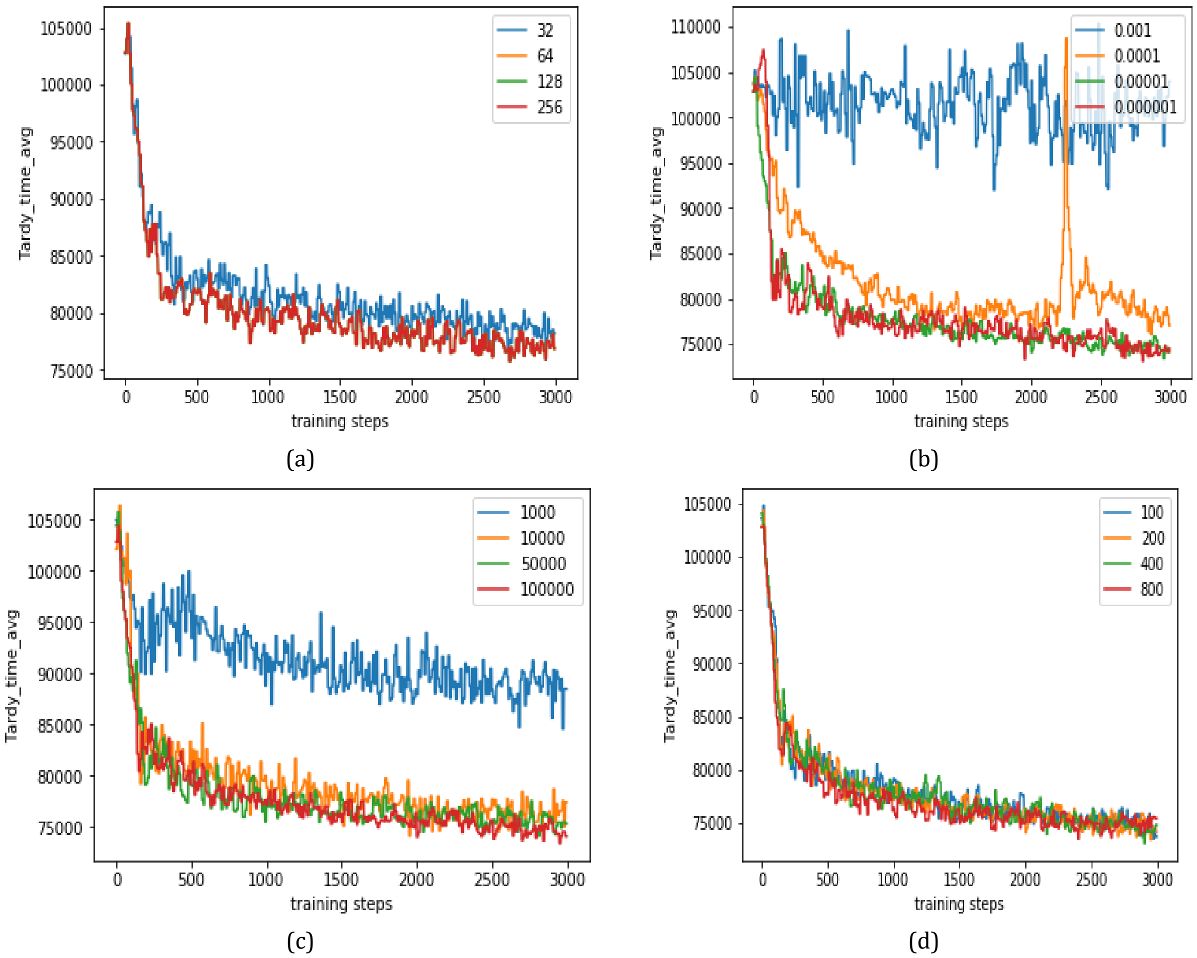


Fig. 3 Verification results of each hyperparameter: (a) Minibatch size; (b) Learning rate; (c) Replay memory buffer size (d) Target network updating frequency

Fig. 3(a) verifies the influence of the training batch on the algorithm. The figure reveals that all parameters exhibits excellent stability. The performance with a batch size of 32 decreased slightly. Fig. 3(b) displays the influence of various learning rates on the algorithm. The higher the learning rate is, the more the training effect is unstable. When the learning rate is 0.001, the algorithm does not even converge. Fig. 3(c) displays the influence of various replay memory buffer sizes on the algorithm. As displayed in the figure, the larger the replay memory is, the better the convergence of the algorithm is, and the replay memory with a capacity of 100000 exhibits superior stability. Fig. 3(d) displays the influence of the target network parameter updating frequency on algorithm performance, and the parameters exhibit an effect under various updating frequencies. According to the verification of various parameters, the final neural network parameter settings are presented in Table 2.

Table 2 Setting of neural network parameters

Parameters	Values
Number of episodes	3000
Explore times steps	3000 · total operation number · 0.3
Epsilon	$1-(1-\varepsilon_{min}\cdot\min{(1,\text{current}_{\text{iter}}/\text{totalsteps}))}$
Replay memory buffer size	100000
Learning rate	0.000001
Minibatch	256
Target network updating frequency	200
Discount factor	1.0

The ε-greedy action selection strategy was implemented according to the method mentioned in a previous study [33]. To ensure the maximum cumulative rewards learned by the agent, the discount factor is selected as 1.0, that is, the cumulative rewards are not discounted.

4.2 Comparison of various status features expression

To verify the validity of the state feature expression of the proposed five-channel images, the influence of three state feature expression modes of three-channel images, four-channel images, and five-channel images on the algorithm were compared. In three-channel images, the production system state feature expression method in literature is adopted [33]. Five-channel images were the proposed production system state feature expression method, and four-channel images were separated from five-channel images to remove the waiting time channel of each operation in the waiting queue.

Five groups of data were randomly generated for testing under the following production configurations: the number of machines was 10, the average time interval between two consecutive new order arrival was 50, DDT was 1.5, and the number of new order insertion were 10, 30, and 50. The results are displayed in Fig. 4. The figure reveals that with the increase in the new order insertion scale, the expression methods of three-channel and four-channel both fluctuated considerably, whereas the proposed expression method exhibited excellent stability.

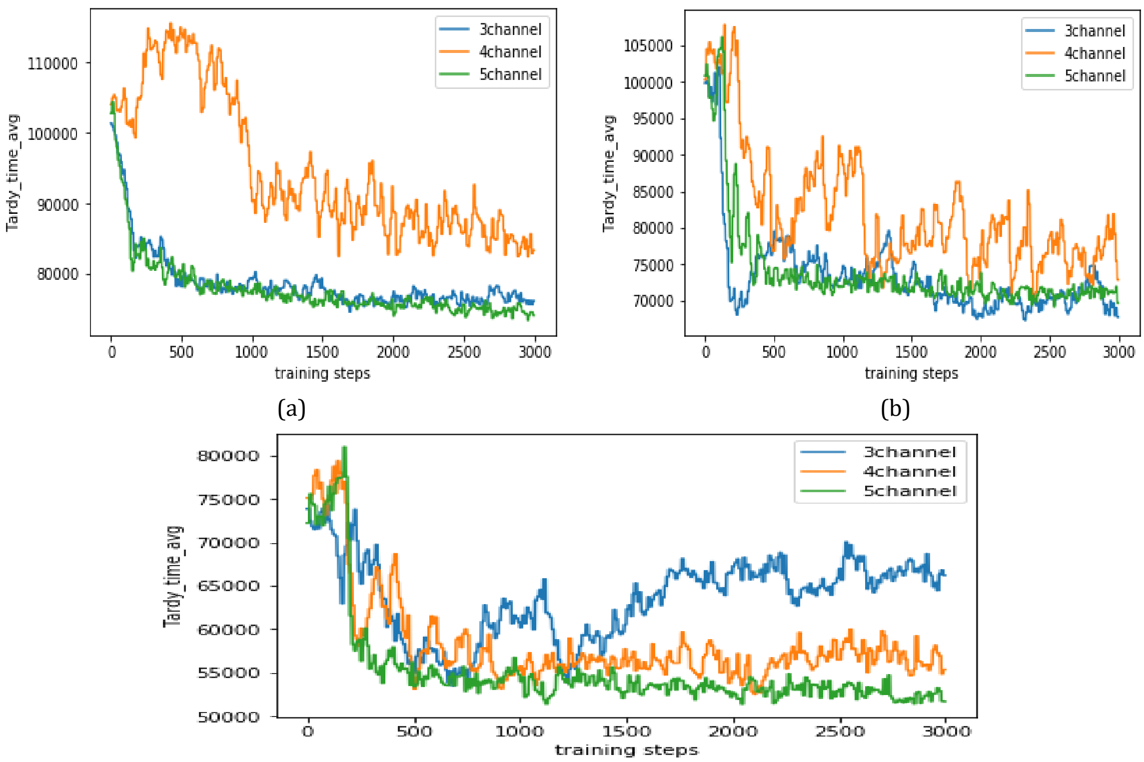


Fig. 4 Verification results of various state features: (a) 10 new job insertions; (b) 30 new job insertions; (c) 50 new order insertions

4.3 Training process of the DDDQN

The DDDQN was used to train a model with certain generalization ability. The model was tested with various test data. The model was trained according to the number of machine and the time interval between the dynamic arrival of orders. Each model was trained for 3000 episodes. In the training process, 12 groups of randomly generated data were used according to the number of new orders of 10, 30, and 50 and the DDT of 1.0 and 2.0. Fig. 5 displays the model training process with five machines and 100 times intervals.

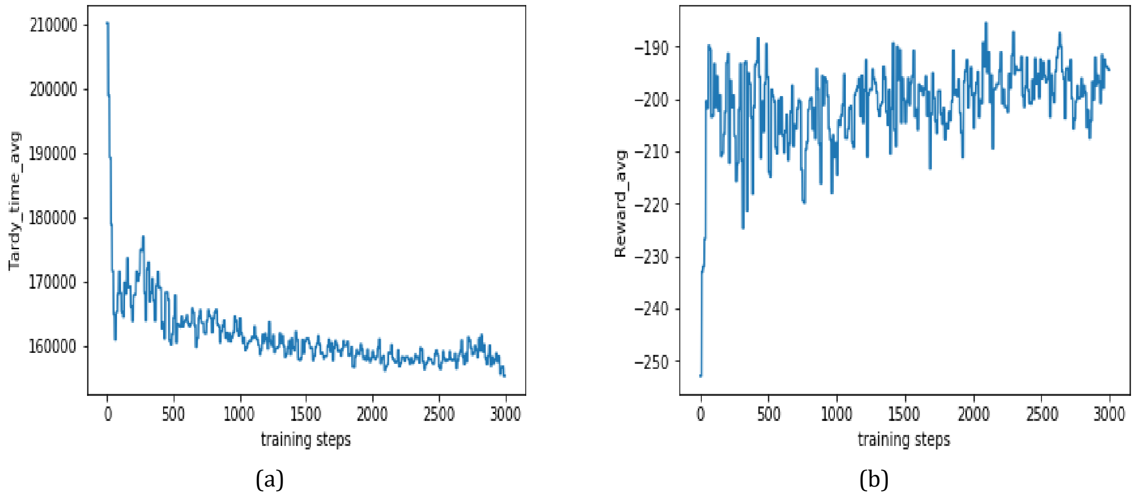


Fig. 5 Model training process in five machines and 100 times intervals: (a) Average total tardiness during training; (b) Average reward during training

Figs. 5(a) and 5(b) display the change process of the total tardiness and reward during the training process. The reward value gradually increases and becomes stable with the increase in training episodes, whereas the total tardiness decreases. After 1000 episodes of training, the model becomes stable, which indicates that the DDDQN has learned to adaptively select the appropriate dispatching rules at the decision time. The curve trend of the average reward value is similar to that of the average total tardiness, which indicates that the designed reward function exhibited a high correlation with the optimization objective with the minimum total tardiness. A small fluctuation was observed after the model convergence. The fluctuation was related to the exploration mechanism of DRL.

4.4 Comparison with conventional dispatching rules

To verify whether the training model can select appropriate dispatching rules at various decision moments, 16 dispatching rules were compared on the test data set. Test data were configured for 81 production scenarios according to all parameter configurations in Table 1, and 30 groups of test data were randomly generated for each production scenario. Tables 3-5 displays the comparison results under various number of machines. The data in the table are the average values of test data. The results of the optimal values are displayed in bold for easy identification. The test results indicate that the algorithm model of the DDDQN is superior to the single scheduling rule in most cases, which reveals satisfactory solution solving ability and generalization ability of various problems. Finding a scheduling rule that can perform well in all production scenarios is difficult.

Table 3 Test results compared with dispatching rules under five machines

Eavg	DDT	n_{add}	DDDQN	SPT	LPT	LWKR	MWKR	SSO	LSO	SRM	LRM	FIFO	EDD	SPT+ SSO	LPT+ LSO	SPT/ TWK	LPT/ TWK	SPT* TWK	LPT* TWK
		10	**17357**	19686	28070	17510	29855	20858	27870	18753	29631	27863	18483	18352	27245	22350	27779	17884	27068
	1.0	30	**32305**	38132	57816	32469	64009	40996	60024	35495	64420	52818	36364	34790	56965	44306	57974	33911	56239
		50	**45832**	53785	87842	45887	101770	57570	96690	50743	103294	74163	52510	48775	87674	64105	87011	48203	85888
		10	**14908**	17218	25612	15059	27386	18393	25401	16299	27162	25395	15896	15887	24780	19883	25310	15417	24614
25	1.5	30	**28547**	34363	54069	28717	60239	37230	56253	31743	60648	49046	32326	31024	53201	40542	54208	30146	52510
		50	**40911**	48853	82932	40980	96842	52642	91754	45841	98358	69226	47351	43851	82756	59182	82083	43278	81013
		10	**12571**	14771	23215	12742	24920	15972	22935	13974	24694	22926	13377	13487	22342	17472	22855	13006	22266
	2.0	30	**24932**	30656	50447	25145	56486	33541	52504	28151	56878	45275	28323	27369	49518	36926	50492	26516	48963
		50	**36174**	44031	78168	36321	91953	47829	86824	41178	93428	64290	42214	39115	77991	54527	77227	38562	76427
		10	**16838**	19093	27552	16932	29539	20199	27498	18072	29347	26825	17965	17791	26742	21664	27282	17274	26818
	1.0	30	**22246**	24766	39989	22278	43158	27871	39063	24663	42527	34269	24550	23414	38877	28719	39383	22724	39588
		50	26119	27639	49293	26601	51064	32367	45594	30007	49455	38772	28549	26413	45654	32019	48605	**25648**	47967
		10	**14346**	16584	25049	14444	27028	17692	24988	15572	26834	24313	15400	15284	24240	19155	24773	14765	24329
50	1.5	30	**18721**	21317	36542	18822	39756	24381	35655	21185	39110	30782	21034	19946	35451	25304	35889	19270	36192
		50	21876	23515	44928	22292	47016	28009	41471	25624	45358	34521	24043	22165	41481	27941	44191	**21458**	43722
		10	**12003**	14118	22644	12113	24533	15235	22492	13222	24329	21800	12904	12865	21800	16748	22298	12355	21981
	2.0	30	**15568**	18155	33328	15700	36565	21106	32420	17997	35857	27380	17561	16755	32305	22259	32545	16166	33112
		50	18074	19912	41027	18597	43405	24099	37736	21828	41625	30466	19865	18475	37867	24497	40117	**17910**	40067
		10	**15173**	17021	25108	15352	26651	18225	24509	16480	26259	23709	16125	16014	24234	19305	24820	15536	24473
	1.0	30	**15748**	17485	27199	15902	27681	19793	25101	17467	27181	24300	16815	16495	25888	19741	27063	15994	26604
		50	**14768**	16025	25094	14954	25826	17951	23477	16130	25365	22904	15595	15418	23825	18441	24543	14841	24458
		10	**12705**	14556	22641	12893	24196	15747	22049	14013	23799	21234	13584	13548	21785	16846	22330	13070	22025
100	1.5	30	**12998**	14804	24421	13163	25045	17017	22444	14700	24530	21577	14014	13789	23212	17087	24245	13304	23876
		50	**12189**	13498	22442	12376	23327	15326	20982	13507	22852	20335	13043	12847	21304	15935	21861	12311	21896
		10	**10417**	12202	20294	10641	21823	13362	19656	11744	21393	18780	11157	11218	19466	14562	19914	10746	19771
	2.0	30	**10560**	12375	21850	10756	22603	14459	19976	12243	22066	18967	11374	11343	20777	14748	21616	10924	21456
		50	**9874**	11166	20028	10097	20997	12896	18637	11209	20480	17860	10602	10512	19004	13700	19346	10034	19620

Table 4 Test results compared with dispatching rules under ten machines

Eavg	DDT	nadd	DDDQN	SPT	LPT	LWKR	MWKR	SSO	LSO	SRM	LRM	FIFO	EDD	SPT+ SSO	LPT+ LSO	SPT/ TWK	LPT/ TWK	SPT* TWK	LPT* TWK
		10	20981	22591	31577	21132	32255	25542	28239	21785	31198	30193	21617	22329	30241	24314	31177	**20934**	30617
	1.0	30	**33733**	37593	55837	33878	59908	43261	52166	35861	57985	50718	36140	37020	54543	40785	55596	34486	54841
		50	**54353**	60774	95044	54712	105769	71725	88968	57266	102888	81281	58799	60474	90305	66574	93949	55778	92471
		10	15896	17492	26486	16055	27155	20447	23143	16704	26098	25092	16393	17230	25154	19218	26081	**15834**	25540
25	1.5	30	**26682**	30549	48797	26854	52863	36216	45126	28843	50938	43670	28891	29974	47514	33743	48551	27442	47835
		50	**44665**	51057	85352	45018	96054	62014	79251	47592	93170	71562	48740	50757	80611	56868	84235	46064	82813
		10	11218	12496	21544	11469	22067	15486	18173	12142	21006	19992	11478	12280	20249	14388	21066	**10899**	20735
	2.0	30	**20163**	23680	41964	20383	45856	29296	38215	22321	43910	36623	21926	23159	40714	27071	41575	20637	41182
		50	**35443**	41592	75924	35946	86383	52458	69664	38515	83493	61844	38873	41317	71231	47648	74616	36716	73633
		10	**19524**	20891	29663	19760	30786	23937	26574	20428	29779	27904	20322	20955	28548	22412	29925	19604	29158
	1.0	30	**26583**	28375	45016	27452	47407	33446	39682	28857	44996	37856	28725	28281	43301	30408	45299	26766	44594
		50	33767	35203	61152	35939	60039	44526	49236	37475	56292	47583	37539	36658	56320	37714	61198	**33730**	60224
		10	**14441**	15795	24575	14687	25691	18845	21486	15359	24682	22804	15189	15856	23472	17331	24826	14507	24112
50	1.5	30	**19467**	21245	37726	20195	40338	26211	32554	21585	37958	30614	21423	21100	36089	23334	37989	19633	37378
		50	24179	25669	51299	26152	50594	34705	39658	27685	46865	37768	27502	27022	46608	28274	51316	**24152**	50496
		10	9738	10907	19714	10149	20688	13936	16602	10799	19671	17706	10285	11003	18614	12662	19854	**9662**	19438
	2.0	30	**13470**	14995	31111	13963	34026	19701	26216	15344	31656	23763	14672	14816	29650	17493	31154	13625	30927
		50	544470	18053	42552	18184	42803	26247	31776	19565	39129	29031	18947	19225	38292	21088	42314	16826	42168
		10	**17211**	18062	26805	17587	27139	21091	23701	18277	26008	24274	17847	18353	25360	19446	26445	17240	26114
	1.0	30	17389	18268	27961	18330	27786	21831	23995	18980	26194	24260	18605	18548	26906	19710	27501	**17384**	27545
		50	19839	20904	31334	21012	31206	24680	27157	21590	29455	27729	21248	21469	29566	22318	31428	**19836**	30832
		10	**12234**	13112	21796	12588	22226	16078	18769	13282	21107	19283	12781	13378	20406	14518	21417	12275	21164
100	1.5	30	**11913**	12824	22216	12632	22401	16191	18572	13283	20822	18697	12879	13048	21299	14331	21721	11922	21848
		50	**13601**	14726	24832	14437	25137	18308	20932	15070	23434	21397	14678	15240	23231	16205	24887	13601	24408
		10	**7743**	8658	17171	8316	17646	11477	14229	8998	16517	14436	8156	8897	15876	10249	16670	7859	16747
	2.0	30	**7413**	8379	17390	8232	17813	11454	14046	8774	16273	13799	8194	8571	16635	10075	16779	7527	17197
		50	**8800**	9903	19595	9633	20152	13151	15982	10256	18472	16105	9647	10369	18204	11612	19556	8910	19386

Table 5 Test results compared with dispatching rules under fifteen machines

Eavg	DDT	nadd	DDDQN	SPT	LPT	LWKR	MWKR	SSO	LSO	SRM	LRM	FIFO	EDD	SPT+ SSO	LPT+ LSO	SPT/ TWK	LPT/ TWK	SPT* TWK	LPT* TWK
		10	22058	23233	31924	22220	31724	26677	27884	22971	30298	29570	22884	23507	30281	24643	31552	**22022**	31409
	1.0	30	**43428**	47477	66167	43666	69721	54727	58929	44931	67134	60692	45554	48028	62792	50423	65351	44820	65288
		50	**62010**	69072	101619	62266	112907	80190	90506	64825	109394	88322	65971	68955	96448	74621	100463	64080	100294
		10	14403	15539	24249	14606	24024	18982	20203	15343	22598	21870	15239	15812	22619	16960	23866	**14331**	23779
25	1.5	30	**31929**	35961	54661	32194	58207	43217	47418	33470	55619	49176	34300	36519	51295	38915	53842	33308	53851
		50	**46976**	54008	86584	47275	97842	65126	75445	49835	94329	73255	50662	53895	81410	59562	85398	49019	85297
		10	7334	8288	17063	8336	16362	11655	12974	8995	14954	14210	8249	8596	15475	10082	16485	**7202**	16848
	2.0	30	**21257**	24694	43555	21926	46752	31911	36133	23125	44153	37669	23234	25351	40249	27998	42562	22244	43059
		50	**32904**	39352	71957	33690	82912	50335	60800	36150	79406	58199	36263	39264	66871	45302	70549	34677	70861
		10	**21029**	21865	29949	21193	30426	25171	26476	21726	28820	27873	21855	22078	28617	23108	29644	21066	29784
	1.0	30	**32289**	33714	51706	32983	52959	40450	43539	34230	49804	43101	34524	34815	48151	35389	51248	32480	50473
		50	38373	38744	63575	40050	62705	47647	51403	41284	59086	49754	41548	40156	58561	40858	61844	**38085**	62948
		10	**13358**	14180	22297	13574	22754	17491	18809	14122	21152	20183	14013	14396	20978	15452	21965	13396	22172
50	1.5	30	**21044**	22477	40327	21716	41791	29076	32281	22884	38709	31722	23277	23554	36876	24233	39815	21236	39162
		50	23960	24485	48881	25429	48478	33044	37038	26681	44917	35094	26630	25716	43984	26674	47048	**23727**	48349
		10	**6461**	7219	15262	7335	15351	10340	11753	7849	13776	12608	7349	7381	13961	8828	14691	6494	15355
	2.0	30	**11824**	13065	30074	12674	31974	18937	22485	13555	29025	21046	12782	14019	26907	15283	29286	12116	29269
		50	**13149**	13821	36133	14338	37025	20928	25650	15344	33588	22410	14083	14595	31933	16432	33995	13226	36082
		10	18289	19347	27279	19016	26979	22250	23649	19559	25844	24482	19284	19640	25772	20125	26824	**18281**	26956
	1.0	30	**20760**	21456	32037	21560	30585	25180	26591	22381	28830	27015	22220	22138	29728	22390	31795	20778	31263
		50	**22665**	23284	35045	23668	32977	27470	29123	24374	31529	29369	24431	23971	33033	24490	34917	22666	34470
		10	10811	11865	19689	11512	19569	14712	16171	12039	18434	16913	11776	12159	18248	12704	19215	**10803**	19424
100	1.5	30	**11766**	12475	22497	12301	21643	15950	17502	13102	19956	17813	12990	13113	20475	13539	22283	11809	21898
		50	**12626**	13301	24121	13092	22945	17011	18764	13839	21531	18970	13640	13940	22381	14583	23942	12638	23689
		10	**4589**	5622	12967	5758	12852	8112	9721	6253	11736	9797	5394	5876	11688	6840	12424	4707	13028
	2.0	30	**5302**	5996	15253	6139	14699	8970	10770	6872	13072	10376	6079	6604	13564	7505	14796	5426	14896
		50	**5889**	6585	16380	6687	15730	9626	11608	7302	14393	11133	6549	7092	15003	8084	16061	6069	16318

4.5 Comparison with the GA algorithm

To prove the computational speed and optimization ability of the model, the DDDQN was compared with the GA. In the GA, an active decoding approach and an elite retention strategy are used. For the medium- and large-scale problems, the GA first generates an initial scheduling scheme based on the initial order data. When a new order arrives, the GA reschedules to generate a new scheduling scheme. The start processing time of all orders differs considerably, and the number of the remaining operations of each order also differs. The parameters of the GA are set as follows: population size is 50, the crossover rate is 0.9, the mutation rate is 0.1, and the iteration number is 300. Some representative data were selected for verification. Each group of test data consists of 30 randomly generated data. The results are presented in Table 6. The data in the table are the average values of test data. The scheduling results and calculation time of the model are superior to the GA in all cases. The average calculation time of the DDDQN model to generate the scheduling scheme for test data at each decision moment was 0.05 s, which was almost instantaneous. Thus, the model can be used for real-time scheduling.

Table 6 Comparison results of DDDQN and GA

m	Eavg	DDT	nadd	Total tardiness		CPU times (s)	
				DDDQN	GA	DDDQN	GA
5	25	1.0	10	17357	26263	0.04	46.29
			50	45832	64650	0.02	57.38
		2.0	10	12571	21354	0.04	46.26
			50	36174	53352	0.02	56.34
	100	1.0	10	15173	22436	0.04	35.42
			50	14768	23195	0.01	8.11
		2.0	10	10417	18107	0.04	35.82
			50	9874	17173	0.01	8.10
10	25	1.0	10	20981	30552	0.10	92.18
			50	54353	107956	0.05	163.86
		2.0	10	11218	20531	0.10	91.82
			50	35443	88138	0.05	164.19
	100	1.0	10	17211	28167	0.10	82.69
			50	19839	36521	0.04	28.24
		2.0	10	7743	18470	0.10	82.35
			50	8800	22720	0.04	28.30
15	25	1.0	10	22058	32677	0.18	139.78
			50	62010	130030	0.08	266.73
		2.0	10	7334	17217	0.15	138.98
			50	32904	98901	0.08	265.63
	100	1.0	10	18289	31864	0.18	131.80
			50	22665	48197	0.08	57.58
		2.0	10	4589	17439	0.16	131.53
			50	5889	26732	0.07	58.57

5. Conclusion

A DRL algorithm, namely the DDDQN, was proposed to solve real-time dynamic job shop scheduling with new order insertions. SPP technology was applied to the neural network structure. A five-channel production system state feature expression method that considered both global and local feature information was considered. As the action space, 16 commonly used dispatching rules were used, and the corresponding reward function was designed to minimize total tardiness. Finally, considerable data from various production scenarios were generated at random to train and test the system model.

Compared with conventional dispatching rules and heuristic algorithms, the results revealed that the algorithm outperformed the single scheduling rule method in most cases, which indicated that the algorithm can select dispatching rules adaptively in various production states. Compared with the GA, the computational speed and optimization ability of the trained models were validated, and real-time optimization and online decision were performed in dynamic event disturbance.

In the future, numerous uncertain factors, such as emergency orders, order cancellations, uncertain processing times, equipment failures, and other multiple disturbance factors, will be studied. Compared with the pure full connection layer neural network, the CNN exhibits a complex structure, which renders model training speed slow. The DQN in this study is a value-based method that cannot directly optimize the policy. Therefore, policy-based DRL methods, such as A3C and PPO, should be studied to improve the quality of solutions and the training speed.

References

[1] Zhou, J., Li, P.G., Zhou, Y.H., Wang, B.C., Zang, J.Y., Meng, L. (2018). Toward new-generation intelligent manufacturing, *Engineering*, Vol. 4, No. 1, 11-20, doi: 10.1016/j.eng.2018.01.002.

[2] Wang, X.H., Duan, H.B. (2014). A hybrid biogeography-based optimization algorithm for job shop scheduling problem, *Computers & Industrial Engineering*, Vol. 73, 96-114, doi: 10.1016/j.cie.2014.04.006.

[3] Çaliş, B., Bulkan, S. (2015). A research survey: Review of AI solution strategies of job shop scheduling problem, *Journal of Intelligent Manufacturing*, Vol. 26, No. 5, 961-973, doi: 10.1007/s10845-013-0837-8.

[4] Baykasoğlu, A., Karaslan, F.S. (2017). Solving comprehensive dynamic job shop scheduling problem by using a GRASP-based approach, *International Journal of Production Research*, Vol. 55, No. 11, 3308-3325, doi: 10.1080/00207543.2017.1306134.

[5] Bokrantz, J., Skoogh, A., Ylipää, T., Stahre, J. (2016). Handling of production disturbances in the manufacturing industry, *Journal of Manufacturing Technology Management*, Vol. 27, No. 8, 1054-1075, doi: 10.1108/JMTM-02-2016-0023.

[6] Zhang, F.F., Mei, Y., Nguyen, S., Zhang, M.J. (2020). Evolving scheduling heuristics via genetic programming with feature selection in dynamic flexible job-shop scheduling, *IEEE Transactions on Cybernetics*, Vol. 51, No. 4, 1797-1811, doi: 10.1109/TCYB.2020.3024849.

[7] Kundakcı, N., Kulak, O. (2016). Hybrid genetic algorithms for minimizing makespan in dynamic job shop scheduling problem, *Computers & Industrial Engineering*, Vol. 96, 31-51, doi: 10.1016/j.cie.2016.03.011.

[8] Cao, H.J., Zhou, J., Jiang, P., Hon, K.K.B., Yi, H., Dong, C.Y. (2020). An integrated processing energy modeling and optimization of automated robotic polishing system, *Robotics and Computer-Integrated Manufacturing,* Vol. 65, Article No. 101973, doi: 10.1016/j.rcim.2020.101973.

[9] Ning, T., Huang, M., Liang, X., Jin, H. (2016). A novel dynamic scheduling strategy for solving flexible job-shop problems, *Journal of Ambient Intelligence and Humanized Computing*, Vol. 7, No. 5, 721-729, doi: 10.1007/s12652-016-0370-7.

[10] Fan, W., Zheng, L.Y., Ji, W., Xu, X., Lu, Y.Q., Wang, L.H. (2021). A machining accuracy informed adaptive positioning method for finish machining of assembly interfaces of large-scale aircraft components, *Robotics and Computer-Integrated Manufacturing*, Vol. 67, Article No. 102021, doi: 10.1016/j.rcim.2020.102021.

[11] Zhang, S.C., Wong, T.N. (2017). Flexible job-shop scheduling/rescheduling in dynamic environment: A hybrid MAS/ACO approach, *International Journal of Production Research*, Vol. 55, No. 11, 3173-3196, doi: 10.1080/00207543.2016.1267414.

[12] Park, S.C., Raman, N., Shaw, M.J. (1997). Adaptive scheduling in dynamic flexible manufacturing systems: A dynamic rule selection approach, *IEEE Transactions on Robotics and Automation*, Vol. 13, No. 4, 486-502, doi: 10.1109/70.611301.

[13] Wang, Z., Zhang, J.H., Yang, S.X. (2019). An improved particle swarm optimization algorithm for dynamic job shop scheduling problems with random job arrivals, *Swarm and Evolutionary Computation,* Vol. 51, Article No. 100594, doi: 10.1016/j.swevo.2019.100594.

[14] Caldeira, R.H., Gnanavelbabu, A., Vaidyanathan, T. (2020). An effective backtracking search algorithm for multi-objective flexible job shop scheduling considering new job arrivals and energy consumption, *Computers & Industrial Engineering,* Vol. 149, Article No. 106863, doi: 10.1016/j.cie.2020.106863.

[15] Ghaleb, M., Zolfagharinia, H., Taghipour, S. (2020). Real-time production scheduling in the Industry-4.0 context: Addressing uncertainties in job arrivals and machine breakdowns, *Computers & Operations Research,* Vol. 123, Article No. 105031, doi: 10.1016/j.cor.2020.105031.

[16] Tang, D.B., Dai, M., Salido, M.A., Giret, A. (2016). Energy-efficient dynamic scheduling for a flexible flow shop using an improved particle swarm optimization, *Computers in Industry*, Vol. 81, 82-95, doi: 10.1016/j.compind.2015.10.001.

[17] Panwalkar, S.S., Iskander, W. (1977). A survey of scheduling rules, *Operations Research*, Vol. 25, No. 1, 45-61, doi: 10.1287/opre.25.1.45.

[18] Lu, M.-S., Romanowski, R. (2013). Multicontextual dispatching rules for job shops with dynamic job arrival, *International Journal of Advanced Manufacturing Technology*, Vol. 67, 19-33, doi: 10.1007/s00170-013-4765-8.

[19] Zhang, H., Roy, U. (2019). A semantics-based dispatching rule selection approach for job shop scheduling, *Journal of Intelligent Manufacturing*, Vol. 30, No. 7, 2759-2779, doi: 10.1007/s10845-018-1421-z.

[20] Zhang, F.F., Mei, Y., Zhang, M.J. (2019). A new representation in genetic programming for evolving dispatching rules for dynamic flexible job shop scheduling, In: Liefooghe, A., Paquete, L. (eds.), *Evolutionary Computation in Combinatorial Optimization. EvoCOP 2019. Lecture Notes in Computer Science,* Vol 11452. Springer, Cham, Switzerland, 33-49, doi: 10.1007/978-3-030-16711-0_3.
[21] Ferreira, C., Figueira, G., Amorim, P. (2022). Effective and interpretable dispatching rules for dynamic job shops via guided empirical learning, *Omega*, Vol. 111, Article No. 102643, doi: 10.1016/j.omega.2022.102643.
[22] Kaelbling, L.P., Littman, M.L., Moore, A.W. (1996). Reinforcement learning: A survey, *Journal of Artificial Intelligence Research*, Vol. 4, 237-285, doi: 10.1613/jair.301.
[23] Sutton, R.S., Barto, A.G. (2018). *Reinforcement learning: An introduction*, Second edition, MIT press, Cambridge, Massachusetts, USA.
[24] Wang, Y.-C., Usher, J.M. (2004). Learning policies for single machine job dispatching, *Robotics and Computer-Integrated Manufacturing,* Vol. 20, No. 6, 553-562, doi: 10.1016/j.rcim.2004.07.003.
[25] Chen, X.L., Hao, X.C., Lin, H.W., Murata, T. (2010). Rule driven multi objective dynamic scheduling by data envelopment analysis and reinforcement learning, In*: Proceedings of 2010 IEEE International Conference on Automation and Logistics,* Hong Kong, China, 396-401, doi: 10.1109/ICAL.2010.5585316.
[26] Qu, S.H., Wang, J., Shivani, G. (2016). Learning adaptive dispatching rules for a manufacturing process system by using reinforcement learning approach, In: *Proceedings of 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, Berlin, Germany, 1-8, doi: 10.1109/ETFA.2016.7733712.
[27] Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A. (2017). Deep reinforcement learning: A brief survey, *IEEE Signal Processing Magazine*, Vol. 34, No. 6, 26-38, doi: 10.1109/MSP.2017.2743240.
[28] Zhu, J., Wang, H., Zhang, T. (2020). A deep reinforcement learning approach to the flexible flowshop scheduling problem with makespan minimization, In: *Proceedings of 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)*, Liuzhou, China, 1220-1225, doi: 10.1109/DDCLS49620.2020.9275080.
[29] Luo, S. (2020). Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning, *Applied Soft Computing*, Vol. 91, Article No. 106208, doi: 10.1016/j.asoc.2020.106208.
[30] Yang, S., Xu, Z., Wang, J. (2021). Intelligent decision-making of scheduling for dynamic permutation flowshop via deep reinforcement learning, *Sensors*, Vol. 21, No. 3, Article No. 1019, doi: 10.3390/s21031019.
[31] Li, Y.X., Gu, W.B., Yuan, M.H., Tang, Y.M. (2022). Real-time data-driven dynamic scheduling for flexible job shop with insufficient transportation resources using hybrid deep Q network, *Robotics and Computer-Integrated Manufacturing*, Vol. 74, Article No. 102283, doi: 10.1016/j.rcim.2021.102283.
[32] Liu, C.-L., Chang, C.-C., Tseng, C.-J. (2020). Actor-critic deep reinforcement learning for solving job shop scheduling problems, *IEEE Access*, Vol. 8, 71752-71762, doi: 10.1109/ACCESS.2020.2987820.
[33] Han, B.-A., Yang, J.-J. (2020). Research on adaptive job shop scheduling problems based on dueling double DQN, *IEEE Access*, Vol. 8, 186474-186495, doi: 10.1109/ACCESS.2020.3029868.
[34] Wang, L.B., Hu, X., Wang, Y., Xu, S.J., Ma, S.J., Yang, K.X., Liu, Z.J., Wang, W.D. (2021). Dynamic job-shop scheduling in smart manufacturing using deep reinforcement learning, *Computer Networks*, Vol. 190, Article No. 107969, doi: 10.1016/j.comnet.2021.107969.
[35] He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 9, 1904-1916, doi: 10.1109/TPAMI.2015.2389824.

APEM journal
Advances in Production Engineering & Management **ISSN 1854-6250**
Volume 18 | Number 2 | June 2023 | pp 152–162 Journal home: apem-journal.org
https://doi.org/10.14743/apem2023.2.463 Original scientific paper

Genetic algorithm-based approach for makespan minimization in a flow shop with queue time limits and skipping jobs

Han, J.H.[a], **Lee, J.Y.**[b,*]

[a]Department of Industrial Engineering, Pusan National University, Republic of Korea
[b]Division of Business Administration & Accounting, Kangwon National University, Republic of Korea

ABSTRACT

This study investigates a flow shop scheduling problem with queue time limits and skipping jobs, which are common scheduling requirements for semiconductor and printed circuit board manufacturing systems. These manufacturing systems involve the most complex processes, which are strictly controlled and constrained to manufacture high-quality products and satisfy dynamic customer orders. Further, queue times between consecutive stages are limited. Given that the queue times are limited, jobs must begin the next step within the maximum queue time after the jobs in the previous step are completed. In the considered flow shop, several jobs can skip the first step, referred to as skipping jobs. Skipping jobs exist because of multiple types of products processed in the same flow shop. For the considered flow shop, this paper proposes a mathematical programming formulation and a genetic algorithm to minimize the makespan. The GA demonstrated its strengths through comprehensive computational experiments, demonstrating its effectiveness and efficiency. As the problem size increased, the GA's performance improved noticeably, while maintaining acceptable computation times for real-world fab facilities. We also validated its performance in various scenarios involving queue time limits and skipping jobs, to further emphasize its capabilities.

ARTICLE INFO

Keywords:
Scheduling;
Flow shop;
Makespan;
Queue time limits;
Skipping jobs;
Optimization;
Modeling;
Genetic algorithm

**Corresponding author:*
Jy.lee@kangwon.ac.kr
(Lee, J.Y.)

Article history:
Received 9 October 2022
Revised 4 July 2023
Accepted 6 July 2023

1. Introduction

This study focused on the scheduling problem in a three-machine flow shop with queue time limits and skipping jobs. Such scheduling problems are found in semiconductor and printed circuit board (PCB) manufacturing systems, which play key roles in the electronics industry as the major components of most electronic products. Such manufacturing systems generally fabricate multiple types of products. Therefore, meeting customer demand in terms of quality and improving the throughput rate are the most important objectives. Queue time limits and skipping jobs are related to product quality and multiple types of products, respectively. In this paper, we investigate the scheduling problem in the considered flow shop with the objective of minimizing makespan subject to queue time limits and skipping jobs.

Semiconductor wafer fabrication is generally a time-critical production environment owing to many highly reactive chemical processes. Thus, the queue times of work-in-process wafers in a fab are strictly managed and constrained to satisfy wafer quality requirements [1]. These constraints are common in wafer lot scheduling, especially at Korean semiconductor manufacturing companies that enforce queue time limits across multiple process steps to enhance overall quality.

Generally, semiconductor manufacturing involves limited queue times for two main reasons. Firstly, it is crucial to maintain the cleanliness of the wafer surfaces during the waiting period before proceeding to the next stage. Prolonged exposure of wafer surfaces to air heightens the risk of contamination by impurities, leading to significant quality issues. Secondly, it is essential to preserve the chemical efficacy treated in the previous step while awaiting the subsequent stage. Therefore, chemically treated wafers must proceed to the next step within a specified timeframe known as the limited queue time. Failure to initiate the subsequent processing step within this queue time limit may require reworking or discarding the wafers, especially in cases of severe contamination [2]. Consequently, this causes lower productivity and business losses. As semiconductor processes become more complex, the number of processes with queue time limits increases, increasing their importance to semiconductor manufacturing. In addition to semiconductor manufacturing systems, queue time limits can also be found in many manufacturing systems for batteries, food, steel, and crude oil [33].

In a semiconductor manufacturing fab, several types of wafer products are produced simultaneously. Their main process flows are similar, but some specific steps are different depending on the product types. Thus, flow shops modeled in a semiconductor fab are designed to process multiple product types, some skipping unnecessary steps. For example, if the first step is cleaning or metrology, some lots that do not need this step skip it and go straight to the second step. For such jobs, the queue time limits are set between the second and third steps. With an increase in the number of semiconductor products, this feature has become more common. Flow shops with skipping jobs are also found in PCB manufacturing systems for multi-variety low-volume production. Additionally, skipping jobs are found in various manufacturing systems, including pharmaceuticals, molten iron, stainless steel, and flour [1, 27, 30].

A typical flow shop consists of a series of machines or workstations arranged in sequential order and processes every job through the machines in that order. If the orders of all jobs are the same across all machines, this is referred to as a permutation schedule. In general, when job i completes on machine k but job j is still unfinished on machine $(k + 1)$, job i can wait until job j completes. However, in a flow shop with queue time limits, job i cannot wait for a long time. In other words, after job i completes on machine k, its subsequent operation on machine $(k + 1)$ must begin within a specified limited queue time. On the other hand, Most flow shop scheduling studies have assumed that every job processes at all stages, which is currently not valid because of the increased customization and diversification of products [3].

Using three-field notation in [4] to define theoretically the considered scheduling problem, it is $F_3|max\text{-}wait, skip|C_{max}$. The first field means a flow shop with three machines. The second field represents problem characteristics, i.e., queue time limits and skipping jobs, respectively. The last field shows the makespan as the objective function to be minimized. As mentioned in [5], a two-machine flow shop with queue time limits is NP-hard, so this scheduling problem is also NP-hard.

We provide a mixed-integer programming (MIP) formulation to describe the scheduling problem clearly. The MIP can be solved using a commercial optimization solver. However, obtaining optimal solutions using the solver may require a significant computation time or it cannot be achieved within an acceptable time, given that the problem is NP-hard. Therefore, we propose a genetic algorithm to rapidly obtain adequate and effective solutions. The genetic algorithm was evaluated under different cases of computational experiments, and it showed highly effective and efficient performance.

The structure of this paper is as follows. Section 2 outlines the assumptions and notation used to describe the proposed algorithms and the mixed-integer programming formulation for the scheduling problem. Section 3 introduces heuristic algorithms, while their evaluation is discussed in Section 4. Lastly, Section 5 provides the study's conclusions and a summary of the findings.

2. Literature survey

With reference to the available literature, most flow shop scheduling studies with queue time limits (QTL) have focused on two-machine problems. Yang *et al*. [5] proved that minimizing the makespan in a two-machine flow shop with QTL is NP-hard and proposed a branch-and-bound (B&B) algorithm. Additionally, advanced lower bounds and dominance properties were developed in [6, 7], while a constructive heuristic algorithm was proposed in [8]. Furthermore, An *et al*. [9] included not only QTL but also sequence-dependent setup times and developed heuristic algorithms and a B&B algorithm, while Lee [10] suggested a genetic algorithm to minimize the total tardiness for the same flow shop. In addition, Dhouib *et al*. [11] proposed a MIP and a simulated annealing algorithm to hierarchically minimize the number of tardy jobs and makespan for multi-machine flow shops with QTL, whereas Kim and Lee [12] introduced a three-machine flow shop with overlapping QTL. To minimize the total tardiness, Hamdi and Loukil [13] suggested a lower bound scheme based on Lagrangian relaxation and a heuristic algorithm, and Joo *et al*. [14] performed simulation experiments with a list scheduling rule. Furthermore, flow shop problems with QTL have been considered in several recent studies [15-23].

In the relevant literature, scheduling problems with skipping jobs have not been extensively investigated in comparison with other types of flow shops. However, a feature of skipping jobs has recently received significant research attention, given that multiple product types with different specifications are produced in the same manufacturing fabrication. Rajendran and Ziegler [24] examined the performance of dispatching rules and a heuristic algorithm to minimize total flow time. Saravanan *et al*. [25] and Dios *et al*. [26] proposed a simulated annealing algorithm and various heuristic algorithms, respectively, to minimize the makespan, whereas, Saravanan *et al*. [27] suggested a genetic algorithm to minimize the mean tardiness. On the other hand, Tseng *et al*. [28] proposed a heuristic to change a given permutation schedule to an improved non-permutation schedule for minimizing the makespan, while Li *et al*. [29] proposed a multi-objective artificial bee colony algorithm with respect to flowtime, earliness, and tardiness in molten iron processing. Recently, in the context of Industry 4.0 production, Rossit *et al*. [30] introduced a flow shop with skipping jobs.

In the literature, QTL and skipping jobs in flow shop scheduling have been separately studied, and their incorporation of them has received limited attention, with only a few studies exploring them together. Notably, for no-wait flow shop problems with skipping jobs, Glass *et al*. [31] proposed and analyzed several heuristic algorithms for minimizing the makespan in a two-machine, while Smutnicki *et al*. [32] devised a method to determine the cyclic schedule with minimal cycle time. For the general queue time limits, Ruiz *et al*. [33] investigated a comprehensive flow shop problem encompassing various features, such as queue time limits, skipping jobs, and the inclusion of machine eligibility, machine release dates, precedence constraints, and sequence-dependent setup times, and they evaluated simple dispatching rules and heuristic algorithms. Additionally, Yu *et al*. [1] focused on a two-machine flow shop with QTL and skipping jobs and they analyzed mathematical properties, explored the reduction of the search space, and developed efficient approximation algorithms for minimizing queue time variations. Furthermore, Han and Lee [34] dealt with minimizing total tardiness in a flow shop scheduling problem that involved queue time limits and skipping jobs.

In the existing literature, there is a notable gap in flow shop research regarding $F_3|max\text{-}wait, skip|C_{max}$. It is important to highlight that the study conducted by Han and Lee [34] focused on minimizing total tardiness in a flow shop scheduling problem specifically within the semiconductor foundry business. This industry places significant emphasis on satisfying customer delivery dates. However, in our present study, we shift our attention to the memory business, where factors such as equipment utilization and throughput take precedence. Thus, the objective of this study is to minimize the makespan. It is worth mentioning that this study represents the first attempt to tackle this particular problem, and the outcomes obtained can serve as a foundational basis for future research and advancements in this field.

3. Problem description

In this section, we describe the considered scheduling problem with assumptions, notation, and MIP formulation. The considered flow shop has three stages (k = 1, 2, 3), and each stage has a machine m_k. All jobs (i = 1, 2,…, n) are given and are available to start at time zero, and all information for making schedules are provided in advance. That is, for job i, processing time PT_{ik} on machine k and queue time limit QT_{ik} between m_k and $m_{(k+1)}$ are already known. Among the jobs, there are jobs starting at m_1, called *normal jobs*, and skipping jobs starting at m_2. For this scheduling problem, the objective is to minimize the makespan which is equal to the completion time of the last processed job on m_3.

To address this problem, we restrict a solution space only to permutation schedules. In particular, given a specific job sequence, all jobs follow that sequence and are processed on machines in the same order. Although an optimal schedule may not be included in the permutation schedules, many studies with QTL have assumed permutation schedules. This is because lots in semiconductor manufacturing fabs are typically processed in permutation schedules for traceability, manageability, and flexibility in material handling [7]. The followings are additional assumptions made in this study.

Pre-emption is not permitted, meaning a job cannot be interrupted once it starts processing. Each job can be processed on only one machine at a time, and each machine can process only one job at a time. The queue time limits must be satisfied without failure. However, the number of jobs queueing between machines is not limited.

Additionally, the following symbols in Table 1 are used to describe the MIP and the proposed algorithms.

Table 1 The symbols used in the algorithms

Symbol	Meaning
h	index of the position in a sequence, h = 1,…, n
$[h]$	index of a job placed at the h-th position in a sequence
x_{ih}	= 1 if job i is placed at the h-th position in a sequence; otherwise, 0
$ST_{[h]k}$	the time at which the h-th job starts on m_k
$CT_{[h]k}$	the time at which the h-th job completes on m_k

We provide equations for the completion times of all jobs in a given sequence. For the first scheduled job (h = 1), Eqs. 1 to 3 calculate the completion times on the three machines, respectively. Obviously, the first job is processed immediately without waiting.

$$CT_{[1]1} = PT_{[1]1} \tag{1}$$

$$CT_{[1]2} = CT_{[1]1} + PT_{[1]2} \tag{2}$$

$$CT_{[1]3} = CT_{[1]2} + PT_{[1]3} \tag{3}$$

From the second position onwards (h = 2,…, n), the calculations for the completion times on the first two machines (m_1 and m_2) differ between the normal jobs starting at m_1 and skipping jobs starting at m_2. Eqs. 4 and 5 computes the completion times of the normal jobs on m_1 and m_2, while Eqs. 6 and 7 computes those of the skipping jobs. The completion times of the final machine are computed using Eq. 8.

$$CT_{[h]1} = max\{CT_{[h-1]1} + PT_{[h]1}, CT_{[h-1]2} - QT_{[h]1}, CT_{[h-1]3} - QT_{[h]2} - PT_{[h]2} - QT_{[h]1}\} \tag{4}$$

$$CT_{[h]2} = max\{max\{CT_{[h]1}, CT_{[h-1]2}\} + PT_{[h]2}, CT_{[h-1]3} - QT_{[h]2}\} \tag{5}$$

$$CT_{[h]1} = CT_{[h-1]1} \tag{6}$$

$$CT_{[h]2} = max\{CT_{[h-1]2} + PT_{[h]2}, CT_{[h-1]3} - QT_{[h]2}\} \tag{7}$$

$$CT_{[h]3} = max\{CT_{[h]2}, CT_{[h-1]3}\} + PT_{[h]3} \tag{8}$$

The next is the MIP formulation to clearly define the considered scheduling problem and to serve a benchmark schedule using an optimization solver. Herein, normal jobs and skipping jobs are not distinguished. In particular, it is assumed that skipping jobs have zero processing time at m1 and sufficiently long queue times between m_1 and m_2, to render the queue time limits non-significant, i.e., $PT_{i1} = 0$ and $QT_{i1} = \infty$ for skipping jobs.

$$\text{[P] Minimize } CT_{[n]3} \tag{9}$$

$$\text{subject to} \sum_i x_{ih} = 1, \quad 1 \le h \le n \tag{10}$$

$$\sum_h x_{ih} = 1, \quad \forall i \tag{11}$$

$$ST_{[h]k} + \sum_i PT_{ik} x_{ih} \le ST_{[h+1]k}, \quad 1 \le h \le n-1, 1 \le k \le 3 \tag{12}$$

$$ST_{[h]k} + \sum_i PT_{ik} x_{ih} \le ST_{[h]k+1}, \quad 1 \le h \le n, 1 \le k \le 2 \tag{13}$$

$$ST_{[h]k} + \sum_i (PT_{ik} + QT_{ik}) x_{ih} \ge ST_{[h]k+1}, \quad 1 \le h \le n, 1 \le k \le 2 \tag{14}$$

$$ST_{[h]3} + \sum_i PT_{i3} x_{ih} \le CT_{[h]3}, \quad 1 \le h \le n \tag{15}$$

$$ST_{[h]k} \ge 0, \quad 1 \le h \le n, 1 \le k \le 3 \tag{16}$$

$$x_{ih} \in \{0, 1\}, \quad \forall i, 1 \le h \le n \tag{17}$$

Eq. 9 represents the objective function's makespan, which is the maximum completion time of the jobs. To ensure permutation schedules without preemption, Eqs. 10 and 11 stipulate that each position can accommodate only one job and each job can and must only occupy one position in the sequence, respectively. Eq. 12 defines the relationship between the start times of two consecutive jobs, while Eq. 13 defines the relationship between the start times on two consecutive machines. To satisfy queue time limits, Eq. 14 guarantees that each job's subsequent operation starts within the specified queue time. Eq. 15 calculates the completion times of the jobs, and Eqs. 16 and 17 establish the decision variables' domains.

4. Used Methods: Genetic algorithms

Because the problem considered is NP-hard, an effective and efficient heuristic algorithm is required to quickly obtain adequate solutions. Thus, we propose a genetic algorithm (GA), which is a nature-inspired evolutionary optimization method. The GA is one of the most commonly used methods because it has low mathematical requirements and is highly flexible in application [35-38]. The following subsections describe the proposed GA design.

4.1 Solution representation and fitness

The solutions to the problem considered are expressed in permutation schedules, which are sequences of job indices that can be directly represented in the chromosome structure of the GA. Each solution has an objective function value, which is the makespan of the schedule and should be minimized. Thus, the makespan to be minimized was used for fitness evaluation.

4.2 Initial population

To generate an initial population, a constructive heuristic algorithm referred to as the NEH (Nawaz, Enscore, and Ham) algorithm was used. Since the NEH algorithm has shown good performance in many flow shop scheduling problems, but it does not guarantee an optimal solution. Thus, it has been commonly used to find an initial solution for further improvement using other optimization techniques.

The NEH is based on the concept of inserting jobs into a sequence iteratively to minimize the makespan. Initially, it calculates the total processing time for each job by summing up the processing times across all machines. Next, the jobs are sorted in decreasing order based on their total processing times. For each job in the sorted order, it is inserted at all possible positions in the sequence, and the resulting makespan is evaluated. Then, the position that yields the lowest makespan is chosen, and the job is inserted at that position. Herein, for population diversity, we used shortest processing time (SPT)-based list scheduling rules to obtain different solutions. The remaining initial population was randomly generated.

- SPT_1 with p_{i1}
- SPT_2 with p_{i2}
- SPT_3 with p_{i3}
- SPT_4 with $p_{i2} + p_{i3}$
- SPT_5 with $p_{i1} + p_{i2} + p_{i3}$

4.3 Selection

The selection operation randomly selects chromosomes for the next generation. The chromosomes are selected with a selection probability proportional to their fitness values. This operation is analogous to the survival of the fittest in the theory of evolution and is referred to as the *roulette* method in the GA. Note that the performance of the roulette method has been demonstrated in recent related studies [14, 21, 22]. Because the objective function should be minimized, the inverse f_i of the objective function value for each chromosome i in the population is computed, and the selection probability $prob_i$ of each chromosome is obtained by $prob_i = f_i / \sum_j f_j$. The next population is composed based on the selection probabilities of the chromosomes.

4.4 Crossover

The crossover operation generates offspring by exchanging information about the selected parents. Various methods can be used for crossover, such as the one-point crossover and two-point crossover methods [14]. Hosseinabadi *et al.* [39] and Hasançebi and Erbatur [40] evaluated various crossover methods for genetic algorithms and found that a one-point crossover demonstrated high performance in scheduling problems. Lee [14] reported that a one-point crossover demonstrated relatively high performance in a flow shop with queue time limits. Based on the existing research results, a one-point crossover was applied in this study.

For each chromosome in the current population, if a randomly generated number (between 0 and 1) is less than the probability of crossover p_c, the chromosome is copied into a pool. They are then randomly paired and referred to as *parents*. For the parents ($Parent_1$ and $Parent_2$), a one-point crossover is applied to generate two offsprings ($Offspring_1$ and $Offspring_2$) using the following procedure:

Step 0: Randomly select a crossover point from among the parent genes.
Step 1: $Offspring_1$ inherits genes up to the crossover point within $Parent_1$.
Step 2: Except for the genes that $Offspring_1$ already contains, $Offspring_1$ inherits genes from $Parent_2$ in the order in which they appear in $Parent_2$.
Step 3: $Offspring_2$ is generated in the same manner in Steps 1 and 2.

4.5 Mutation

The mutation is an operation that prevents premature convergence and maintains the diversity within a population. This operation partially mutates several genes in a chromosome with low probability.

For each chromosome in the population, if a random number (between 0 and 1) is less than the mutation probability, the chromosome is mutated once. The mutated ones are also included in the population. In this study, two common methods were used, namely, *insertion* and *exchange*, with the same probability. Insertion selects a job at random and inserts it into a randomly selected position, whereas exchange interchanges two randomly selected jobs

4.6 Local search

We used a local search technique to improve the performance of the GA. To prevent a rapid increase in the calculation time owing to the local search, it was only applicable to 10 % of the solutions in a population at each generation. In this study, *insertion* and *exchange* methods were used. In particular, *insertion* or *exchange* with the same probability was applied 3n times to the solution, where n is the number of jobs. If the new solution was superior to the current solution, the current solution was replaced.

4.7 Entire procedure

The proposed GA terminates when the maximum number of generations is reached; the overall flowchart is shown in Fig. 1.

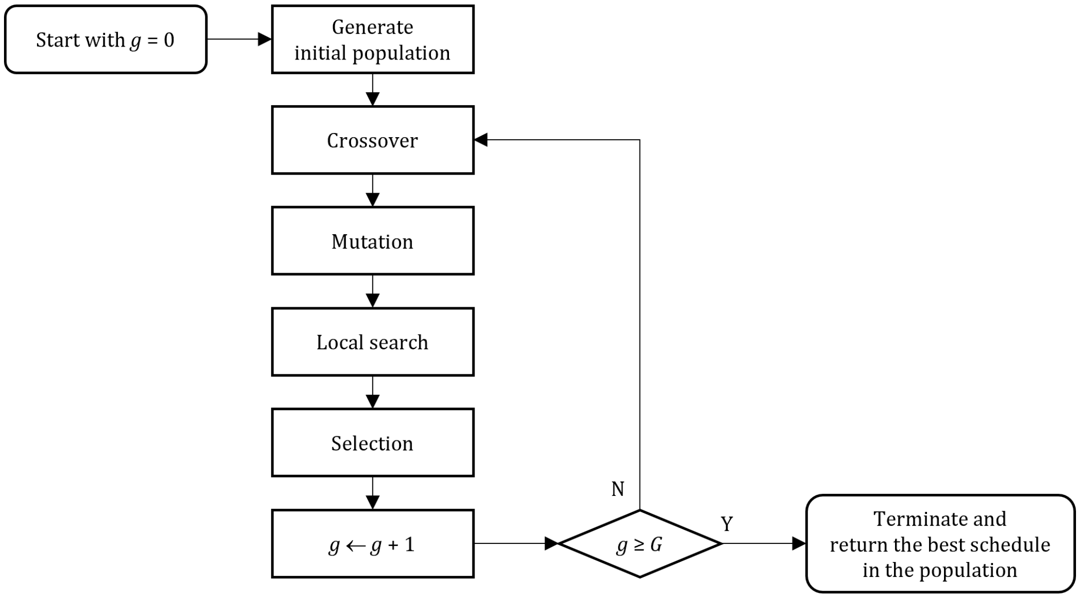


Fig. 1 Flowchart of the proposed GA

5. Computational experiments

To assess the performance of the proposed algorithms, we performed computational experiments using instances generated as follows. The experiments encompassed various levels of the number of jobs, n = (10, 20, 30, 40, 50, 100, 150, and 200). The processing times p were generated from discrete uniform distributions within the range of [1, 50]. Additionally, the queue times were generated from three different distributions with a range of [1, w], where w = (30, 50, and 70). Regarding the proportion of skipping jobs, we considered three levels of the proportion λ of skipping jobs, where λ = (0.3, 0.5, and 0.7), i.e., (0.3n, 0.5n, and 0.7n). For each combination of (n, w, and λ), ten instances were generated, resulting in a comprehensive set of experimental data.

We coded the proposed algorithms using the Java programming language and conducted experiments on a personal computer equipped with an Intel Core i7-8700 CPU running at 3.2GHz. To solve the MIP formulation, we used CPLEX 12.10, a commercial solver, with a maximum CPU time of 1,000 seconds to prevent excessive computation time. As for the proposed GA, it terminated once the maximum number of generations was reached and returned the best solution found within the population. For each instance, the GA was run independently 30 times, and a mode value from the 30 runs was used for the evaluation.

Before evaluating the GA, we performed experimental calibration to improve its performance. In the calibration, we assumed n = 100, λ = 0.5, and w from [1, 50], which were the medium levels of the considered problem instances.

The GA contained four parameter types: population size S, number of generations G, probability of crossover p_c, and probability of mutation p_m. If the population size is excessively small, diverse solutions are not available. On the other hand, if the population size is excessively large, the time required for crossover and local search increases. Thus, excessive time is required to obtain a solution. Therefore, it is necessary to select an appropriate population size related to the number of jobs to be considered. In this study, the size of the population was set to

$(\rho \cdot n)$ and a preliminary experiment was conducted for ρ = 1, 2,…, 5. Considering the trade-off between the algorithm performance and calculation time, ρ was set to 4.

The proposed GA used the maximum number of generations G as the termination condition. Therefore, an appropriate number of generations was required. If G is excessively large, a significant amount of time is required to obtain a solution, whereas, if G is excessively small, it is highly probable that a solution close to the optimal solution will not be obtained. We tested {500, 1000, 1500, 2000, 2500, and 3000} for G. Thereafter, 1000 was set as the appropriate number of generations for the preliminary experiments. In particular, preliminary experiments were conducted on {0.3, 0.5, 0.7, and 0.9} for p_c and {0.05, 0.1, 0.2, 0.3, and 0.4} for p_m, which were set to 0.7 and 0.2, respectively. Note that because of the extensive range of potential combinations among the four types of parameters examined and the resulting substantial amount of data, this paper does not include the results of the preliminary tests.

First, we tested the MIP performance using CPLEX. Table 2 lists the average CPU times of CPLEX with a computation time limit (1,000 s). As seen from the table, CPLEX shows a tendency to work well when the number of jobs skipping the first stage increased and the queue time limits were loose, whereas CPU times increased in the opposite case. The level of difficulty can be estimated based on the problem situation. The CPU time increased significantly as the number of jobs increased. When n = 50, 100, 150, and 200, CLPEX did not obtain an optimal solution for 10, 40, 79, and 89 instances (out of 90) within the time limit.

Next, we evaluated the effectiveness of the simple heuristics used to generate the initial population. Table 3 presents the average percentage error (PE), which is defined as 100 × (obj_A – obj_{CPLEX}) / obj_{CPLEX} for Algorithm A, where N in the names of the last six algorithms represents the NEH algorithm. As seen from the table, the six list scheduling rules provided solutions with large errors compared with those from CPLEX. However, the proposed NEH algorithm demonstrated high performance. Every NEH algorithm obtained schedules with errors of approximately 2 % or less. Moreover, for instances with n = 150 and 200, schedules superior to those obtained from CPLEX were obtained. In addition, all NEH algorithms required less than 1 s. Thus, it was confirmed that an effective and efficient heuristic algorithm should be developed.

In addition, we evaluated the performance of the proposed GA. In addition, to verify the effectiveness of the local search, a GA without a local search (GA_{l-}) was tested. Table 4 presents the performance in comparison with CPLEX, where NI and CPUT indicate the number of instances in which the GA and GA_{l-} obtained a superior schedule to that obtained from CPLEX and the average computation time of the GAs, respectively. The relative deviation index (RDI) is defined as (obj_A – obj_{min}) / (obj_{max} – obj_{min}), where obj_A is the objective value for Algorithm A, and obj_{min} and obj_{max} are the minimum and maximum objective values, respectively. As shown in the table, the effectiveness of the GA was demonstrated more clearly as the number of jobs increased. For all instances with n = 150 and 200, the GA obtained a solution that was superior to that obtained by CPLEX. The performance of the GA was improved using local search. Although the calculation time increased owing to the inclusion of the local search, the average CPUT was 87 s for n = 200, which is practicable.

Table 2 CPU times of CPLEX

λ	w	n=10	20	30	40	50	100	150	200
0.3	30	0.08	231.62	378.25	277.15	688.18	984.75	1000	1000
	50	0.06	0.53	3.95	144.09	313.47	547.89	1000	1000
	70	0.05	0.14	0.31	3.31	39.78	571	940.44	1000
0.5	30	0.06	64.17	4.9	169.36	140.77	818.78	1000	1000
	50	0.04	0.25	3.32	21.56	59.21	432.6	933.92	1000
	70	0.03	0.13	0.89	12.95	12.35	529.04	953.12	1000
0.7	30	0.06	0.46	2.83	111.82	44.83	411.79	852.79	972.94
	50	0.03	0.13	0.89	12.95	12.35	529.04	953.12	1000
	70	0.04	0.09	0.19	10.08	2.62	254.99	804.98	1000

Finally, we evaluated the performance of the GA concerning different problem parameters (λ and w) for instances with n = 100. The results are summarized in Table 5, which presents the average PE values. When both parameters decreased, the PE values also decreased. This indicates that the GA was effective when the limited queue times were strict and the number of jobs that skipped the first stage decreased. Problems with these features are more difficult than others; thus, CPLEX failed to find an optimal or adequate solution to these problems. It should be noted that if λ and w are zero, the problem transforms into a three-machine no-wait flow shop and into a two-machine flow shop problem if both parameters increase.

6. Conclusion

This study investigated a flow shop scheduling problem specifically focusing on the incorporation of queue time limits and skipping jobs, which are critical requirements in semiconductor manufacturing. To address this problem, this study proposed a MIP formulation and GA. Computational experiments were carried out to assess the performance of both the MIP and GA, with a particular emphasis on evaluating the effectiveness and efficiency of the GA. The results demonstrated that as the problem size increased, the GA exhibited improved performance and maintained an acceptable computation time within a real fab facility. Additionally, the GA's performance was verified under various conditions of queue time limits and skipping jobs.

However, it should be noted that in certain conditions during the computational experiments, the GA failed to achieve an optimal solution. Thus, future endeavors should be directed toward enhancing the performance of the GA. Potential approaches for improvement include exploring alternative heuristic algorithms such as state-of-the-art metaheuristics, genetic programming-based algorithms, or machine learning-based algorithms. Furthermore, it is worth considering the integration of bi-objective optimization approaches, incorporating measures based on both makespan and due dates, as the semiconductor business often follows an order-based manufacturing paradigm. Additionally, future research can explore different types of queue time limits,

Table 3 Percent errors of heuristic algorithm

	n=10	20	30	40	50	100	150	200	average
SPT_1	19.23	22.32	21.97	23.99	24.72	25.83	24.58	22.73	23.17
SPT_2	22.41	23.11	22.41	22.99	23.76	24.73	23.64	21.55	23.08
SPT_3	30.00	29.28	27.43	27.21	27.68	27.53	25.59	23.19	27.24
SPT_4	21.91	20.44	18.86	17.84	18.54	17.62	15.92	14.06	18.15
SPT_5	17.05	17.02	14.94	15.54	15.97	16.36	15.12	13.30	15.66
LPT	25.60	22.53	19.93	19.10	18.80	17.72	16.03	14.10	19.23
SPT_1N	1.65	1.23	1.13	1.18	1.19	0.89	-0.54	-2.14	0.57
SPT_2N	2.12	2.00	1.49	1.69	1.48	0.72	-0.76	-2.36	0.80
SPT_3N	2.33	2.03	1.55	1.49	1.37	0.72	-0.86	-2.41	0.78
SPT_4N	1.91	1.33	0.99	0.75	0.64	-0.21	-1.81	-3.35	0.03
SPT_5N	1.50	1.06	0.72	0.81	0.60	-0.08	-1.60	-3.31	-0.04
LPTN	1.31	0.74	0.43	0.35	0.15	-0.43	-1.98	-3.57	-0.37

Table 4 Performance of the proposed GA

n	PE (%)		NI		CPUT (s)		RDI		
	GA	*GA_{I-}*	*GA*	*GA_{I-}*	*GA*	*GA_{I-}*	*GA*	*GA_{I-}*	*CPLEX*
10	0.144	0.281	75	54	0.0	0.0	0.134	0.393	0.000
20	0.106	0.243	81	47	0.2	0.1	0.076	0.476	0.000
30	0.039	0.144	82	54	0.5	0.1	0.052	0.400	0.000
40	0.065	0.129	75	48	1.0	0.3	0.121	0.472	0.022
50	-0.050	0.010	82	57	1.8	0.4	0.068	0.393	0.108
100	-0.554	-0.532	89	76	12.1	1.8	0.011	0.181	0.444
150	-2.071	-2.054	90	90	38.0	4.6	0.000	0.010	0.878
200	-3.612	-3.608	90	90	87.3	9.3	0.000	0.001	0.989

Table 5 Percent errors of the proposed GA on different parameters (n = 100)

w	λ			average
	30	50	70	
0.3	-2.377	-0.499	-0.360	-1.079
0.5	-1.185	-0.261	-0.043	-0.496
0.7	-0.097	-0.157	-0.004	-0.086
average	-1.220	-0.306	-0.136	-0.554

including overlapping queue time limits as presented in [11] and generalized skipping jobs that possess the ability to skip any stage. Lastly, extending the investigation of this problem to hybrid flow shops with multiple machines at each stage could be a valuable direction for future research.

Acknowledgment

This work was supported by the Technology Development Program(RS-2022-00140527) funded by the Ministry of SMEs and Startups(MSS, Korea)

References

[1] Yu, T.-S., Kim, H.-J., Lee, T.-E. (2017). Minimization of waiting time variation in a generalized two-machine flowshop with waiting time constraints and skipping jobs, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 30, No. 2, 155-165, doi: 10.1109/Tsm.2017.2662231.

[2] Lee, J.-H., Zhao, C., Li, J.S., Papadopoulos, C.T. (2018). Analysis, design, and control of Bernoulli production lines with waiting time constraints, *Journal of Manufacturing Systems*, Vol. 46, 208-220, doi: 10.1016/j.jmsy.2018.01.001.

[3] Henneberg, M., Neufeld, J.S. (2016). A constructive algorithm and a simulated annealing approach for solving flowshop problems with missing operations, *International Journal of Production Research*, Vol. 54, No. 12, 3534-3550, doi: 10.1080/00207543.2015.1082670.

[4] Graham, R.L., Lawler, E.L., Lenstra, J.K., Rinnooy-Kan, A.H.G. (1979). Optimization and approximation in deterministic sequencing and scheduling: A survey, *Annals of Discrete Mathematics*, Vol. 5, 287-326, doi: 10.1016/S0167-5060(08)70356-X.

[5] Yang, D.-L., Chern, M.-S. (1995). A two-machine flowshop sequencing problem with limited waiting time constraints, *Computers & Industrial Engineering*, Vol. 28, No. 1, 63-70, doi: 10.1016/0360-8352(94)00026-J.

[6] Bouquard, J.-L., Lenté, C. (2006). Two-machine flow shop scheduling problems with minimal and maximal delays, *4OR*, Vol. 4, 15-28, doi: 10.1007/s10288-005-0069-7.

[7] Joo, B.-J., Kim, Y.-D. (2009). A branch-and-bound algorithm for a two-machine flowshop scheduling problem with limited waiting time constraints, *Journal of the Operational Research Society*, Vol. 60, No. 4, 572-582, doi: 10.1057/palgrave.jors.2602598.

[8] Wang, B.L., Huang, K., Li, T. (2018). Permutation flowshop scheduling with time lag constraints and makespan criterion, *Computers & Industrial Engineering*, Vol. 120, 1-14, doi: 10.1016/j.cie.2018.04.021.

[9] An, Y.-J., Kim, Y.-D., Choi, S.-W. (2016). Minimizing makespan in a two-machine flowshop with a limited waiting time constraint and sequence-dependent setup times, *Computers & Operations Research*, Vol. 71, 127-136, doi: 10.1016/j.cor.2016.01.017.

[10] Dhouib, E., Teghem, J., Loukil, T. (2013). Lexicographic optimization of a permutation flow shop scheduling problem with time lag constraints, *International Transactions in Operational Research*, Vol. 20, No. 2, 213-232, doi: 10.1111/j.1475-3995.2012.00876.x.

[11] Lee, J.-Y. (2020). A genetic algorithm for a two-machine flowshop with a limited waiting time constraint and sequence-dependent setup times, *Mathematical Problems in Engineering*, Vol. 2020, Article ID 8833645, doi: 10.1155/2020/8833645.

[12] Kim, H.-J., Lee, J.-H. (2019). Three-machine flow shop scheduling with overlapping waiting time constraints, *Computers & Operations Research*, Vol. 101, 93-102, doi: 10.1016/j.cor.2018.06.009.

[13] Hamdi, I., Loukil, T. (2015). Minimizing total tardiness in the permutation flowshop scheduling problem with minimal and maximal time lags, *Operational Research*, Vol. 15, No. 1, 95-114, doi: 10.1007/s12351-014-0166-5.

[14] Joo, B.J., Kim, Y.D., Bang, J.Y. (2009). A scheduling algorithm for workstations with limited waiting time constraints in a semiconductor wafer fabrication facility, *Journal of the Korean Institute of Industrial Engineers*, Vol. 35, No. 4, 266-279.

[15] Zhou, N., Wu, M., Zhou, J. (2018). Research on power battery formation production scheduling problem with limited waiting time constraints, In: *Proceedings of 2018 10th International Conference on Communication Software and Networks,* Chengdu, China, 497-501, doi: 10.1109/ICCSN.2018.8488247.

[16] Hamdi, I., Toumi, S. (2019). MILP models and valid inequalities for the two-machine permutation flowshop scheduling problem with minimal time lags, *Journal of Industrial Engineering International*, Vol. 15, 223-229, doi: 10.1007/s40092-019-00331-1.

[17] Samarghandi, H. (2019). Minimizing the makespan in a flow shop environment under minimum and maximum time-lag constraints, *Computers & Industrial Engineering*, Vol. 136, 614-634, doi: 10.1016/j.cie.2019.07.048.

[18] Li, Y., Dai, Z. (2020). A two-stage flow-shop scheduling problem with incompatible job families and limited waiting time, *Engineering Optimization*, Vol. 52, No. 3, 484-506, doi: 10.1080/0305215X.2019.1593974.

[19] Maassen, K., Perez-Gonzalez, P., Gunther, L.C. (2020). Relationship between common objective functions, idle time and waiting time in permutation flow shop scheduling, *Computers & Operations Research*, Vol. 121, Article No. 104965, doi: 10.1016/j.cor.2020.104965.

[20] Jeong, B.J., Han, J.-H., Lee, J.-Y. (2021). Metaheuristics for a flow shop scheduling problem with urgent jobs and limited waiting times, *Algorithms*, Vol. 14, No. 11, Article No. 323, doi: 10.3390/a14110323.

[21] Lee, J.H., Kim, H.J. (2021). 3-machine flow shop scheduling with overlapping waiting time constrains to minimize total completion time, *Journal of the Korean Society of Supply Chain Management*, Vol. 21, No. 3, 1-12, doi: 10.25052/KSCM.2021.12.21.3.1.

[22] Lee, J.H., Yu, T.S., Park, K.S. (2021). Scheduling of flow shop with overlapping waiting time constraints using genetic algorithm, *Journal of the Korean Institute of Industrial Engineers*, Vol. 47, No. 1, 34-44, doi: 10.7232/JKIIE.2021.47.1.034.

[23] Ištoković, D., Perinić, M., Borić, A. (2021). Determining the minimum waiting times in a hybrid flow shop using simulation-optimization approach, *Tehnički Vjesnik – Technical Gazette*, Vol. 28, No. 2, 568-575, doi: 10.17559/TV-20210216132702.

[24] Rajendran, C., Ziegler, H. (2001). A performance analysis of dispatching rules and a heuristic in static flowshops with missing operations of jobs, *European Journal of Operational Research*, Vol. 131, No. 3, 622-634, doi: 10.1016/S0377-2217(00)00105-3.

[25] Saravanan, M., Sridhar, S., Harikannan, N. (2014). Optimization of realistic multi-stage hybrid flow shop scheduling problems with missing operations using meta-heuristics, *International Journal of Engineering Technology*, Vol. 6, No. 1, 484-496.

[26] Dios, M., Fernandez-Viagas, V., Framinan, J.M. (2018). Efficient heuristics for the hybrid flow shop scheduling problem with missing operations, *Computers & Industrial Engineering*, Vol. 115, 88-99, doi: 10.1016/j.cie.2017.10.034.

[27] Saravanan, M., Sridhar, S., Harikannan, N. (2016). Minimization of mean tardiness in hybrid flow shop with missing operations using genetic algorithm, *Journal of Advanced Manufacturing Systems*, Vol. 15, No. 2, 43-55, doi: 10.1142/S0219686716500050.

[28] Tseng, C.-T., Liao, C.-J., Liao, T.-X. (2008). A note on two-stage hybrid flowshop scheduling with missing operations, *Computers & Industrial Engineering*, Vol. 54, No. 3, 695-704, doi: 10.1016/j.cie.2007.09.005.

[29] Li, J.-Q., Pan, Q.-K., Duan, P.-Y. (2016). An improved artificial bee colony algorithm for solving hybrid flexible flowshop with dynamic operation skipping, *IEEE Transactions on Cybernetics*, Vol. 46, No. 6, 1311-1324, doi: 10.1109/Tcyb.2015.2444383.

[30] Rossit, D.A., Toncovich, A., Rossit, D.G., Nesmachnow, S. (2021). Solving a flow shop scheduling problem with missing operations in an Industry 4.0 production environment, *Journal of Project Management*, Vol. 6, No. 1, 33-44, doi: 10.5267/j.jpm.2020.10.001.

[31] Glass, C.A., Gupta, J.N.D., Potts, C.N. (1999). Two-machine no-wait flow shop scheduling with missing operations, *Mathematics of Operations Research*, Vol. 24, No. 4, 911-924, doi: 10.1287/moor.24.4.911.

[32] Smutnicki, C., Pempera, J., Bocewicz, G., Banaszak, Z. (2022). Cyclic flow-shop scheduling with no-wait constraints and missing operations, *European Journal of Operational Research*, Vol. 302, No. 1, 39-49, doi: 10.1016/j.ejor.2021.12.049.

[33] Ruiz, R., Şerifoğlu, F.S., Urlings, T. (2008). Modeling realistic hybrid flexible flowshop scheduling problems, *Computers & Operations Research*, Vol. 35, No. 4, 1151-1175, doi: 10.1016/j.cor.2006.07.014.

[34] Han, J.-H., Lee, J.-Y. (2022). Scheduling for a flow shop with waiting time constraints and missing operations in semiconductor manufacturing, *Engineering Optimization*, doi: 10.1080/0305215X.2022.2124406.

[35] Xu, W., Sun, H.Y., Awaga, A.L., Yan, Y., Cui, Y.J. (2022). Optimization approaches for solving production scheduling problem: A brief overview and a case study for hybrid flow shop using genetic algorithms, *Advances in Production Engineering & Management*, Vol. 17, No. 1, 45-56, doi: 10.14743/apem2022.1.420.

[36] Shi, D.L., Zhang, B.B., Li, Y. (2020). A multi-objective flexible job-shop scheduling model based on fuzzy theory and immune genetic algorithm, *International Journal of Simulation Modelling*, Vol. 19, No. 1, 123-133, doi: 10.2507/IJSIMM19-1-CO1.

[37] Boudjemline, A., Chaudhry, I.A., Rafique, A.F., Elbadawi, I.A., Aichouni, M., Boujelbene, M. (2022). Multi-objective flexible job shop scheduling using genetic algorithms, *Tehnički Vjesnik – Technical Gazette*, Vol. 29, No. 5, 1706-1713, doi. 10.17559/TV-20211022164333.

[38] Chen, D., Zhao, X.R. (2021). Production management of hybrid flow shop based on genetic algorithm, *International Journal of Simulation Modelling*, Vol. 20, No. 3, 571-582, doi: 10.2507/IJSIMM20-3-CO12.

[39] Hosseinabadi, A.A.R., Vahidi, J., Saemi, B., Sangaiah, A.K., Elhoseny, M. (2019). Extended genetic algorithm for solving open-shop scheduling problem, *Soft Computing*, Vol. 23, No. 13, 5099-5116, doi: 10.1007/s00500-018-3177-y.

[40] Hasançebi, O., Erbatur, F. (2000). Evaluation of crossover techniques in genetic algorithm based optimum structural design, *Computers & Structures*, Vol. 78, No. 1-3, 435-448, doi: 10.1016/S0045-7949(00)00089-4.

APEM journal

Advances in Production Engineering & Management
Volume 18 | Number 2 | June 2023 | pp 163–174
https://doi.org/10.14743/apem2023.2.464

ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Hybrid forecasting modelling of cost and time entities for planning and optimizing projects in the die-cast aluminium industry

Muñoz-Ibañez, C.[a], Chairez, I.[a], Jimenez-Martinez, M.[b], Molina, A.[a], Alfaro-Ponce, M.[a,*]

[a]Tecnologico de Monterrey, Institute of Advanced Materials for Sustainable Manufacturing, Mexico City, Mexico
[b]Tecnologico de Monterrey, School of Science and Engineering, Puebla, Mexico

A B S T R A C T

The techniques employed to manage an industrial project are based on tools that aim to achieve the objectives set by an organization. Most of these techniques consider the development of operative and predictive models. The difficulty in developing project planning models relies on estimating large sets of parameters and the need to include model sections of poorly identifiable, that increase costs and time. This work develops a hybrid forecasting model for all the phases that make up die-casting projects through a series of parameters and sub-models that contemplate the particularities of each case, thereby achieving greater precision in the forecast. The model identifies the cost and time factors that affect project planning, specifically in the die-casting industry, and intends to predict their future behaviour when certain initially given conditions are modified. To estimate the parameters of the hybrid model, several factors in the processes were considered that interact in this industry, such as primary matter costs and activities associated to the process. The considered processes that have a substantial economic impact on the implementation of the project were selected. The criteria for this selection considered identifying the relevant parts of the design and manufacturing in the die-casting industry. Process factors such as the Cost of aluminium and its related activities, whose processes will be grouped into cost and time entities to build a set of metrics that allow better control over them. Finally, the proposed model is based on analytical, parametric, and analog methods that achieve accuracy greater than 85 % in predicting the time and Cost of the process.

A R T I C L E   I N F O

Keywords:
Hybrid models;
Entity modelling;
Project planning;
Forecasting models;
Aluminium die-casting;
Cost factors;
Time factors;
Optimization

**Corresponding author:*
marielalfa@gmail.com
(Alfaro-Ponce, M.)

Article history:
Received 5 May 2023
Revised 17 June 2023
Accepted 19 June 2023

1. Introduction

Project planning is useful for improving time and costs in company projects involving management techniques. These approaches have been designed and developed since the middle of the 20th century [1-4] intending to improve the outcome of project planning, the development of qualitative [5-7] and quantitative models [8-11] has ben considered. These models have motivated the implementation several optimization strategies that have contributed to reducing operation costs and operative times of processes included in the project.

Developing valuable models (usually known as estimation models) in project planning requires analyzing historical information regarding the effects and results of different operative strategies and control techniques implemented in previous projects. Due to the specific conditions of each productive sector where the estimation models could be applied, they must depend on different operative, commercial, and economic variables, including the technology for developing the project, raw materials, workforce, etc. These highly specialized models have been developed for sectors as varied as software [12], construction [13-14], aerospace [15, 16], and automotive [17, 18], machining [19], among others. Specifically, the use of models for estimating manufacturing costs and time, widely used to determine the Cost of forged parts [20], rotating parts [21], manufacturing of die-casting molds [22], and preformed parts employing die cuts [23].

The importance of estimation models in manufacturing processes lies in their ability to accurately predict the costs of primary materials and methods and operative times. In addition, the successful models considered specific needs and practical implications that are part of the environment in which the project is developed. Also, on many occasions, the precision and speed in estimating costs and operative times are associated with the final purchase order, which the customer defines.

With the intention to determine goods production times and costs, the manufacturing processes can be calculated using several methods, like the analytical methods that break down manufacturing activities into their elementary parts. However, despite the benefits of these models, their descriptions usually have many parameters. They are very nonlinear and may not consider some problematic aspects of the model.

A different technique for developing the estimation model is based on non-parametric methods that could produce a mathematical representation of productive relationships using the information generated by the industrial sector under study. These non-parametric models can be complemented with analogous methods to classify product indices according to their dimensional and quality characteristics [24]. All these methods have been successfully used in estimating time and costs for diverse projects of various industrial sectors, including metal mechanics.

In the aluminum die-casting industry, the existing estimation models focused on several factors that affect the manufacturing process. An example of such a method is the model developed by Madan that focuses on determining the optimal Cost and time of production, depending on the geometric conditions of the part; through these data, they determine the number of cavities necessary for the die casting dies [25]. On the other hand, the model developed by Sung focuses on simulating the pores within the die-cast process. Specifically, this model considers factors such as trapped air during the die-casting process caused by the design of the die-casting mold and the properties changes by anisotropic composition due to the die-cast process conditions [26].

Srivastava proposes a model to determine the conditions that cause thermal fatigue, and cracking of casting die [27]. The work presented by Tsoukalas produces a model to determine the porosity of $AlSi_9Cu_3$ aluminum used in die casting. The model is supported by genetic algorithms that validate as main variables the die casting temperature, the die casting mold temperature, and the speed of the phases of die casting [28].

The models described above establish connections between operative conditions in the die-casting process. Nonetheless, they are not formally related to the economic aspects of the final product and its delivery time to the final consumer. Nowadays, many companies apply parametric or probabilistic cost estimation methods. The characteristics and rapid implementation are unique elements of these models. However, these models use synthetic data obtained by approximating the total production cost and overall production time and assuming some desired features for the final product.

The product features can be diverse depending on the design requirements and manufacturing activities. The main problem of this production process is the lack of information about the cost structure and the product's manufacturing processes.

As a result, it became difficult for the designer to visualize the necessary modifications that should be applied to the model to reduce costs in the quotation activities. Such a fact makes it complex to assign only a cost value to the product, limiting the transparent negotiation of the Cost and consequently causing a delay with the customer.

The indirect costs are other factors to consider in the product cost structure, materialized by the support activities. Besides, the causal relationship between cost objects (products and services) and consumable resources is difficult to assess. This evaluation can be solved using traceability, which makes cost analysis explicit in a network with its incorporation into products or services, which is difficult to achieve with the traditional cost estimation approach.

The application of a hybrid model allows accurate forecasting of the phases that make up the processes of die-casting projects. Hybrid models have significant advantages in their use, as shown in the following key applications: a) the optimization of manufacturing processes for obtaining fiberglass in the automotive industry [29], b) the improvement of pill production processes in the pharmaceutical sector [30], and c) in the construction of metal structures at high temperatures through process planning [31]. Likewise, other critical applications for the application of hybrid models that have advantages in precise parameterization in industrial activities are, for example, for inventory control [32], electrical energy consumption [33], in the handling of the materials [34] and the costing for the manufacture of turbines [35].

The advantages of using hybrid models for process planning lie in the flexibility of these models and their ability to adapt to a significant number of features that they may have, selecting the model that best suits the characteristics to be predicted [36]. However, it is also essential to consider that the main disadvantage of its use lies in the large number of data that must be used to achieve considerable precision of the model.

This study presents a data-driven hybrid model to asses costs and process time in the die-cast aluminum industry. The model shown in this work uses analytical, parametric, and analogous methods. Among the variables that are considered for the model development are: a) the Cost of the raw material (which is a determining factor in signing long-term contracts with the automotive and aerospace sector), b) the processing time, and c) the Cost of the operations that will be carried out for the manufacture of the piece, including intermediate process such as die-casting, grinding, die-cutting, drilling, shot blasting, and packing, among others. The goal of the modeling process is to generate a projection that allows controlling the project plan of the die-casting process. In addition, the proposed model considers technical and economic aspects essential for developing projects in the die-casting industry. The successful development becomes the model into an auxiliary tool for estimating the costs and delivery times of the customer product. As a result, the problem of estimating the Cost and time in die-casting projects focuses on constructing a model that can be described as an integral formulation. Such completeness is a consequence of considering a series of sub-models that allow its adaptation to the specific characteristics of the casting process to predict. Hence, this work considers three fundamental aspects of the model design:

- The changes in the price of aluminum in international markets.
- The changes in the production process that have more significant impact on the Cost and the time of the project (die casting, grinding, drilling, and packing, among others).
- The model aspects related to the manufacture of the die-casting mould.

This work is organized as follows. Section 2 is the mathematical basis and the details of the construction of the model. The description of the data set used for its test is included in this part. Section 3, the simulation results achieved by the model. Finally, Sections 4 and 5 close the study with the model's accuracy and relevant conclusions describing some problems related to the discussed topic.

2. Model construction

For the development of the model proposed in this work, it was necessary to estimate the following steps:

- Determination of cost and time entities (CTE) divided according to each primary activity.
- Development of CTE for the manufacturing of parts and manufacturing tooling.
- The CTE for the raw material only considers the Cost, not the time, since it is a complementary value chain activity.
- The CTE's information is gathered from various productive projects to carry out the simulations and the relationships between the projects' time, Cost, and weight.

- The weight directly determines the production costs and expenses to complement the Cost Entity.
- The equations for each CTE are formulated according to the information and the simulation.

2.1 Cost and time entity

For die-casting, estimating Cost and time is the crucial combination for predicting how much it will cost and how long it will take to make a certain product or many products through the analysis of the particularities attributed to it [37]. To identify and control these intrinsic particularities of each development, CTE is proposed. Developed for the analysis of the Cost and time of each critical activity, it provides information related to the added value of the project, so the CTE will allow the estimation of the Cost and time between the products, considering the Cost per kilogram and the time of the product manufactured for homogeneity.

The CTE (Fig. 1) is composed of two elements. Its first element is the Cost Entity (CE), whose primary function is controlling each project phase's Cost. The Time Entity (TE) is its second element, which determines the time assigned to each activity to be carried out.

It's essential to define once the homogeneous resources for the project are stable and interdependent. They are stable because of the imputation of these within the productive chain, which is given by $\frac{\beta}{X}\left(\text{i.e.}, \backslash\frac{\beta}{min}, \frac{\beta}{kg}\right)$. Notice that each resource does not need to change depending on the product. These components are interdependent since the resources are consumed in the same proportion by one element in any way the product is used.

Assuming that $R_i = [k: k \in K]$ represents the number of resources consumed by the CTE_i, TE_i or CE_i to realize the i_{th} activity gave by A_i. If $\left[y_i^k(X_i) = X_i\alpha_k\right]$ is equivalent to the number of resources k consumed by carrying out the activity A_i under the condition X_i (where α_k is the coefficient with which resources k are consumed) and if $C_k = y_i^k(X_i)$ and $T_k = y_i^k(X_i)$ is the allocation fee corresponding to the Cost and time of the resource units and units k. Therefore, Eqs. 1 and 2 are basic for the determination of Cost and time:

$$CE = \sum_{k \in R_i} C_k\left(y_i^k(X_i)\right) \tag{1}$$

$$TE = \sum_{k \in R_i} T_k\left(y_i^k(X_i)\right) \tag{2}$$

The definitions for the distribution of the Cost of a resource would be the following:

- Machining $(nb)\alpha_k\left(\frac{h}{nb}\right)$ an imputation fee $\left(\frac{\beta}{h}\right)$.
- $\left((X_i)\alpha_k\right)$ is an imputation fee, determined by Cost based on weight.
- $y_i^k(X_i)$ is an imputation fee, determined by the time.

The α_k is the coefficient with which resource k is consumed, nb is the identification of the Cost for machining. The coefficients to determine the degree to which resources are consumed within the project activities were determined through statistical relationships obtained from 20 projects within the die-casting industry.

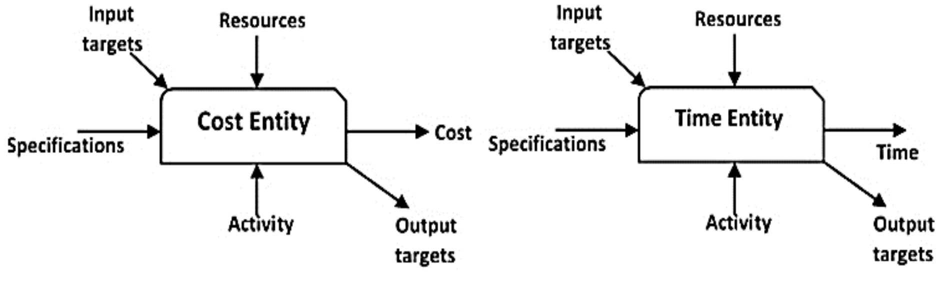


Fig. 1 Graphic representation of the CTE

Cost and time entity pattern

The Cost and Time Entity Pattern (CTEP) is a macro-entity that meets the characteristic of homogeneity of the resources of the CTE. CTEP simplifies the model for determining Cost and time and helps establish an overview of the manufacturing chain.

The basic information for estimating costs and time in projects of the die-casting industry is represented in the Scheme of the Cost and Time Entity Patterns (SCTEP) shown in *supplementary file 1*. The SCTEP is built based on the product's characteristics, such as; physical specifications of the product (weight and type of aluminum), manufacturing (the type of surface finish, drilling, packing, among others), and the tooling (number of cavities, volume, etc.). For the estimation of the Cost and time in the design as well as in the manufacturing, the purchase order is used. Such application allows us to quantify the *flexibility* of the order and give an adequate estimation of the phases necessary to develop the project, so the purchase order represents a priority aspect [38].

It is important to establish that the CE and TE maintain the homogeneity condition understood as the similarity of the activities between the parameters; resources, activities, inputs, and outputs. Therefore, the union of the parameters and the sum of the costs and elementary times of the different cost entity patterns (CEP) and time entity patterns (TEP) are described in Eqs. 3 and 4

$$CEP = \sum_{i \in N} \sum_{k \in R_i} C_k \left(y_i^k (X_i) \right) \tag{3}$$

$$TEP = \sum_{i \in N} \sum_{k \in R_i} T_k \left(y_i^k (X_i) \right) \tag{4}$$

where N is the number of CE that is part of CTEP.

General model for determining cost and time

Once the elements that will be part of the model structure have been defined, it is important to identify the main activities to be modeled. For this work, the activities taken into consideration are:

- The price of the metal.
- The manufacturing time of the product.
- The Cost of the Product.
- The manufacturing time of the die-cast tooling.
- The Cost of the die-cast tooling.

Considering it, the models proposed for the determination of the total Cost and time of the project can be represented as Eqs. 5 and 6.

$$Cost_{Total} = Cost_{RM} + Cost_{manufacturing} + Cost_{Tooling} \tag{5}$$

$$Time_{Total} = Time_{manufacturing} + Time_{Tooling} \tag{6}$$

In some cases, there may be the possibility that the customer no longer requests a part of the project; for example, the customer can provide the tooling. In consequence, it would only be necessary to establish the Cost of the raw material and manufacturing. A binary matrix can be constructed to establish this relation, where *zero* identifies an activity that does not occur and one that occurs in the project.

3. Methodology and primary results

CTEP$_1$ of the raw materials

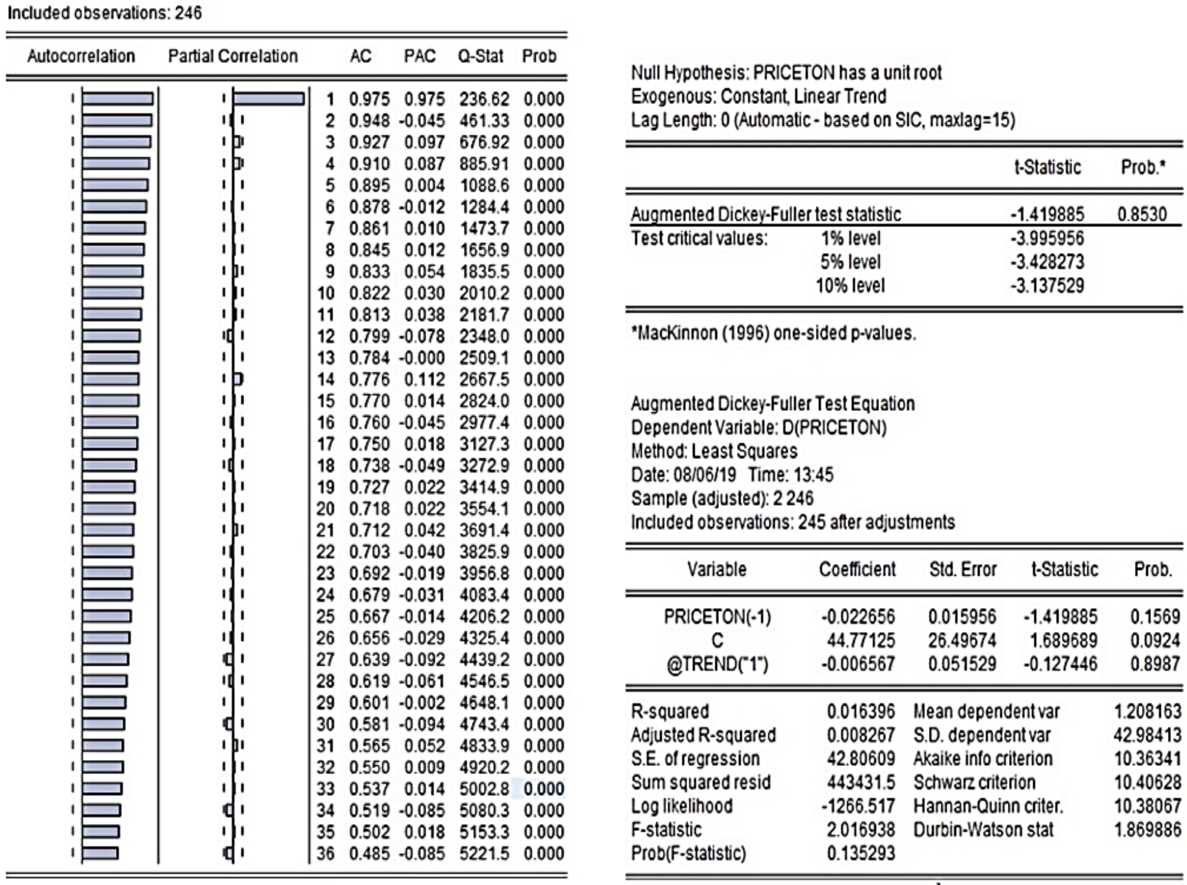
In this case, the CTEP$_1$ for the raw material will include only the calculation of the Cost of aluminum, for which the historical data reported by the London Metal Exchange (LME) of the Cost in dollars per ton will be used. The start date of data collection was January 1, 2016, and the end of it on May 25, 2019. The database contains five monthly readings that will give 246 data for the analysis, this information can be consulted at *supplementary file 2*.

Intending to determine the most accurate forecast model, it is necessary to apply tests to establish the stationarity of the time series, for which the Eq. 7 gives the correlation function of the sample [39].

$$r_k=\left(\sum_{t=1}^{T}(x_t-\bar{x})(x_{t-k}-\bar{x})\right)\left(\sum\nolimits_{t=1}^{T}(x_t-\bar{x})^2\right)^{-1} \tag{7}$$

Therefore, it is possible to develop the correlogram shown in Fig. 2(a), in which the analyzed series of the LME presents a simple and partial first-order correlation, which determines that the immediately previous data are the ones that have the greatest influence on the value of the following data.

$$d=\left(\sum_{t=2}^{T}(e_t-e_{t-1})^2\right)\left(\sum_{t=2}^{T}e_t^2\right)^{-1} \tag{8}$$



Included observations: 246

	AC	PAC	Q-Stat	Prob
1	0.975	0.975	236.62	0.000
2	0.948	-0.045	461.33	0.000
3	0.927	0.097	676.92	0.000
4	0.910	0.087	885.91	0.000
5	0.895	0.004	1088.6	0.000
6	0.878	-0.012	1284.4	0.000
7	0.861	0.010	1473.7	0.000
8	0.845	0.012	1656.9	0.000
9	0.833	0.054	1835.5	0.000
10	0.822	0.030	2010.2	0.000
11	0.813	0.038	2181.7	0.000
12	0.799	-0.078	2348.0	0.000
13	0.784	-0.000	2509.1	0.000
14	0.776	0.112	2667.5	0.000
15	0.770	0.014	2824.0	0.000
16	0.760	-0.045	2977.4	0.000
17	0.750	0.018	3127.3	0.000
18	0.738	-0.049	3272.9	0.000
19	0.727	0.022	3414.9	0.000
20	0.718	0.022	3554.1	0.000
21	0.712	0.042	3691.4	0.000
22	0.703	-0.040	3825.9	0.000
23	0.692	-0.019	3956.8	0.000
24	0.679	-0.031	4083.4	0.000
25	0.667	-0.014	4206.2	0.000
26	0.656	-0.029	4325.4	0.000
27	0.639	-0.092	4439.2	0.000
28	0.619	-0.061	4546.5	0.000
29	0.601	-0.002	4648.1	0.000
30	0.581	-0.094	4743.4	0.000
31	0.565	0.052	4833.9	0.000
32	0.550	0.009	4920.2	0.000
33	0.537	0.014	5002.8	0.000
34	0.519	-0.085	5080.3	0.000
35	0.502	0.018	5153.3	0.000
36	0.485	-0.085	5221.5	0.000

a

Null Hypothesis: PRICETON has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.419885	0.8530
Test critical values:	1% level	-3.995956	
	5% level	-3.428273	
	10% level	-3.137529	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PRICETON)
Method: Least Squares
Date: 08/06/19 Time: 13:45
Sample (adjusted): 2 246
Included observations: 245 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRICETON(-1)	-0.022656	0.015956	-1.419885	0.1569
C	44.77125	26.49674	1.689689	0.0924
@TREND("1")	-0.006567	0.051529	-0.127446	0.8987

R-squared	0.016396	Mean dependent var	1.208163
Adjusted R-squared	0.008267	S.D. dependent var	42.98413
S.E. of regression	42.80609	Akaike info criterion	10.36341
Sum squared resid	443431.5	Schwarz criterion	10.40628
Log likelihood	-1266.517	Hannan-Quinn criter.	10.38067
F-statistic	2.016938	Durbin-Watson stat	1.869886
Prob(F-statistic)	0.135293		

b

Fig. 2 (a) Aluminium price correlogram for LME, and (b) Dickey-Fuller test

Once the correlogram is obtained, the autocorrelation of the data is determined by the statistical Eq. 8 of Durbin-Watson [39], obtaining a value of 1.869886, establishing no autocorrelation.

$$\Delta y_t=\phi_1+\phi_2^t+\alpha y_{t-1}+\sum_{i=1}^{p}\delta_i\Delta_{t-1}+\varepsilon_i \tag{9}$$

On the other hand, the analysis of the stationarity series is carried out using the Dickey-Fuller test as shown in Eq. 9 [39]. The results are shown in Fig. 2(b), setting a decision value of -3.428349 and -1.420186. As a result, it is established that the time series of the LME is not stationary and have a unitary root, so it is possible to model this PEC through an ARIMA model.

With the main characteristics of the time series determined, such as their autocorrelation and non-stationarity, it is possible to select a model that best adapts to the forecast.

In the mentioned above case, an ARMA model (2,2) was chosen, which will be the model for determining the PEC_1; this determines the Cost of the raw material expressed with Eq. 10

$$PEC_1 = Y_{t-1} + \theta(Y_{t-2} - Y_{t-3}) + \varphi(\hat{a}_{t-1}) \tag{10}$$

where the coefficient values for the self-aggressive part expressed by θ is 0.797492 and for the mobile part expressed by φ is 0.833451, and $\alpha_{t-1} = y_{t-1} - \hat{y}_{t-1}$. This model adjust precisely to the series of the Cost of aluminum.

CTEP$_2$ of Product Manufacturing

To build the CTEP$_2$ corresponding to the manufacture of the piece, the first step is to propose the matrix of activities of the processes used in its manufacture so it is possible to establish the parametric relationships. For this, 25 projects were selected from the *Maquinados e Inyecciones Tecamac SA de CV* client portfolio. This company die-casts more than 45 tons of aluminium annually for clients such as Siemens-Mexico and Donaldson, among others, being a second-order supplier of assemblers such as Nissan and GMC. Therefore, the projects (*supplementary file 3*) cover 87 % of the company's annual production. It is important to mention that this Table will become a binary matrix indicating whether the activity is carried out (1) or not (0).

Once the project has defined the activities, the times required by activity for its manufacture are established, and the relationship between these times and the weight of the piece is established. To achieve this, it is necessary to calculate the number of cycles to be observed given by Eq. 11

$$n = \left(\frac{st}{k\hat{X}}\right)^2 \tag{11}$$

where n is the number of cycles to be observed given that s is the standard deviation of the test, k the desired degree of accuracy that is 95 % and $\hat{X}$ the average time, with this a pilot run is established stating that it is necessary to observe 50 cycles. Hence, the study of times of the 25 projects is carried out according to the activities listed in *supplementary file 3*. The study performed for the Optimized Shield in *supplementary file 4* (named as *part name no. 2* in *supplementary file 3*) is taken as an example.

When time and movement studies were carried out for each case, the data fit the best distribution that resembles its behavior. This allows for generating a more significant number of observations artificially. This setting can be seen in *supplementary file 5*, in the case of Optimized Shield time's *part name no. 2*.

This process is repeated for each of the projects, as well as for the operations that are necessary for its manufacture. With this, it is possible to determine the appropriate distribution type, resulting in more than 7,000 data supporting the forecasts; this process is illustrated in *supplementary file 6* for the die-cast operation.

Construction of the TEP$_{2,1}$ for the time of part manufacturing activities

The data shown in *supplementary file 6* allows us to determine the relevant equations for each process and estimate the time for the manufacture of the piece. The resulting equations are shown in *supplementary file 7*, where Y_i^k is a binary variable that takes the value of 1 if the activity is carried out and 0 if the activity is not carried out. Also, X_i is the weight of the piece to be manufactured.

With the specific equations and metrics that govern each of the TE of $TEP_{2,1}$, the general equation is constructed, which is expressed as follows.

$$\begin{aligned} TEP_{2,1} = \sum\nolimits_{k \epsilon R_i} y_i^k ET_i X_i = {} & Y_i^k(27.36 + 51.87X_i) + Y_i^k\left(0.9068 + 60.65X_i - 89.98X_i^2\right) + \\ & Y_i^k(7.94 + 33.47X_i) + Y_i^k(0.78 + 26.20X_i) + Y_i^k(0.61 + 64X_i) \\ & Y_i^k\left(5.15 + 80.63X_i - 163.27X_i^2\right) + Y_i^k(2.22 + 26.33X_i) \end{aligned} \tag{12}$$

Construction of the $CEP_{2,2}$ for the Cost of part manufacturing

For the construction of the CEP, a characterization of the Cost is made concerning kilograms manufactured; for this, the following descriptions were generated:

- *Workforce vs. weight.* For this relationship, the direct, indirect, and administrative workforce is considered as a whole, considering that a relationship is kept according to the kilograms produced by the company.
- *Gas vs. weight.* Gas is the fuel oil used to fuse aluminum in melting furnaces, the second most important raw material. It is directly related to the number of kilograms produced by the company.
- *Electric power vs. weight.* Electric power allows the operation of the die-casting equipment, as well as the various finishing and melting equipment; the consumption of electrical energy is determined directly by the number of kilograms produced by the organization.
- *Expenses vs. weight.* The expenses are all those agents necessary for the organization to carry out its productive activity but are not delivered in the final product; these are distributed equally in the kilograms produced by the company.

The company history is reviewed for the workforce vs. weight evaluation to establish its correspondence; the data is shown in *supplementary file 8*.

Once the workforce's Cost per kilogram is obtained, the next step to determine the CE is to make the corresponding adjustment in the behavior of the data shown in *supplementary file 11*. In this way, the trend will be analyzed, so we can apply a more suitable model for simulation. This procedure is carried out in the same way for the rest of the costs, leaving the equations as shown in Table 1.

Table 1 Cost of the workforce against sales from December 17 to February 19, in Mexican pesos

CE	Equation	Type of distribution
Workforce	$CE_{2,2,1} = N(19.92, 6.40)X_i$	Normal
Gas	$CE_{2,2,2} = N(10.63, 1.38)X_i$	Normal
Power Supply	$CE_{2,2,3} = N(5.77, 3.88, 1.84)X_i$	Triangular
Expenses	$CE_{2,2,4} = N(26.09, 11.89, 7.53)X_i$	Triangular

Where X_i is the weight of the piece to be developed, which will have an allocation rate for the Cost according to the distribution that best fits the item.

Construction of $CTEP_3$ for the Tooling Manufacturing

The complementary activities for the time estimation of machining that help in the tool manufacturing process are the following [40]:

- *Habilitation of the material.* It consists of preparing the steel that will be machined by adjusting the shingles to have a reference cut for the machining; this activity lasts between 5 to 10 percent of the total machining activity.
- *Preparation of tools.* The tools that will be used in the machining operations are selected and prepared; this process lasts 5 to 10 percent of the machining activity.
- *Generation of operations.* The design of the tooling or device is executed according to the specifications and requirements of the client, then the designs are transformed into instructions for the operation of the CNC. The generation of operations is one of the main activities within the manufacture of tooling, and its duration lasts between 30 and 35 percent of machining activities.
- *Machining.* The design is emptied into the material, and a swarf is removed. This activity is carried out with numerical control equipment, which gives precision and certainty. This element gives added value to the manufacturing, so it is the guiding operation within the forecast model.
- *First Mold Assembly.* An initial assembly is carried out to see no interferences inside the cavity and the sealing is adjusted. This operation lasts no more than 2,880 min.

- *Heat treatment.* It is subjected to the cavity at high temperatures to subsequently cool it gradually; this allows the cavity to have a hardness on its surface; this operation lasts no more than 2,880 min compared to the standard established by the company.
- *Second mould assembly.* Once the cavities have been treated thermal, these are assembled again to verify if there are deformations or interferences inside the cavity. If these occur, they are readjusted and returned to verify the closure of the cavity, and this operation does not last more than 2,160 minutes for the standard set by the company.
- *First test run.* The die-cast mould is tested inside the die-cast equipment, allowing them to perform the necessary actions for the final adjustment. This operation does not last more than 1,440 min by the standard established by the company.
- *Final adjustment.* The adjustment of all the observations that were presented in the execution of the test is carried out. In general, these operations include the adjustment of the knobs and the adjustment in the closing of the mould; this operation does not last more than 2,880 minutes, according to the standard established by the company.

Defined the operations that make up the manufacturing of the injection mould, the times obtained from four documented projects are those described in supplementary file 9, where the different machining times are established and likewise, the relationship between the machining is sought, the weight and the number of cavities that make up the tool.

With these data, it is possible to identify the correlation between the weight removed against the weight of the mould (weight of the piece by the number of cavities) by applying the Pearson linear correlation index according to Eq. 13. Which establishes a correlation of 0.9369 and, subsequently, the correlation between the machining time required against the weight eliminated; with this value, it is possible to obtain the $TEP_{2,2}$ as follows:

$$TEP_{3,1} = \sum_{k \epsilon R_i} y_i^k T_k P_i = Y_i^k (705.00 + 161.75 P_i) \tag{13}$$

Where Y_i^k is a binary variable that takes the value of 1 if the activity is carried out, and 0 if the activity is not carried out, P_i which is then withdrawn weight given by $P_i = 0.252 + 155.98 X_i$, and X_i is the weight of the piece to be manufactured. For the estimation of the $ECT_{3,2}$, it is established that it is directly related to the time invested and the fixed and variable costs of the machining activity. The time costs are shown in the Table 2.

When the $CTEP_3$ has been developed for the activity of the manufacture of moulds, it is possible to use the model to forecast the Cost and time within the projects of die casting, as well as to measure its degree of precision.

Table 2 CNC hourly costs for the manufacture of injection molds

Cost per hour	Range in minutes	
	Min	Max
400	0	101663
250	101664	177352
200	177352	∞

4. Model accuracy

To determine the model's accuracy, the test of the χ^2 represented in Eq. 14 will be carried out to establish that the sample obeys the entity model for the estimation of Cost and time. Where V_0 is the real value, and V_e is the expected or predicted value, the k degrees of freedom are adjusted according to the number of observations generated for each test.

$$\chi^2_{(k-1)} = \sum_{k=1}^{N} \frac{(V_0(k) - V_e)^2}{V_e} \tag{14}$$

The test will be applied to each one of the developed sub-models of the CTEP, so the following elements will be evaluated:

- Cost of raw material.
- Cost of the piece.
- Time of manufacture of the piece.
- Tooling cost.
- Tool manufacturing time.

To determine the degree of precision when estimating the Cost of the raw material, a test with 19 degrees of freedom is applied, as shown in *supplementary file 10*. The obtained value of 5.4135 can be interpreted as a 95 % reliability of the model for the Cost of the raw material.

The same procedure follows the rest of the CTEP, which we can summarize in Table 3.

Table 3 Model sensitivity analysis

CTEP	Degree of freedom	χ2	Degree of Reliability (%)	Standard deviation	Kurtosis	Asymmetry	Jarque-Bera
Cost of raw material	19	5.4135	95	1.8213	0.2871	-0.0230	3.6807
Cost of the piece	19	12.1588	85	0.5015	0.0607	-0.0253	4.3209
Time of manufacture of the piece	19	12.9878	80	0.0248	-0.0376	-0.1643	4.6675
Tooling cost	5	2.2916	80	398.44	-1.2481	-0.0665	9.0323.
Tool manufacturing time	5	2.2857	80	0.2059	-1.1548	-0.0580	8.6379

5. Conclusion

The results obtained from the different tests carried out to determine the model's reliability show that it presents a degree of accuracy above 85 percent. This result allows a quick forecast to be made with the minimum information required for the quotation, leading to the construction of an entity-based model to efficiently determine costs and time to provide certainty and a competitive advantage to the die-casting industry.

Analyzing the model structure established that the information necessary for estimating Cost and time is introduced through the different entities built. Such analysis facilitates the person in charge of the quotation's fast and efficient handling of all the data required during the quotation process. These entities of Cost and time have in their main structures the Cost of the raw material, the Cost and time of manufacture of the tools, and the Cost and time of manufacture of the part. Such determination is an advantage in the process since it allows the evaluation of the different projects, although these represent other structures.

An important innovation that the use of CEP provided was that the model increases accuracy and facilitates user interaction. In the same way, it was possible to gather information to generate a solid and well-structured database on which the model can base its approximations.

Finally, in the development of future research, it would be possible to use other modeling systems, such as genetic algorithms or artificial intelligence, to increase the accuracy of the proposed model.

Acknowledgment

The authors would like to thank the Tecnologico de Monterrey seed founding program, projects IDs IJXT070-22TE60001 and IKXT025-22UI60001, and Consejo Mexiquense de Ciencia y Tecnología grant "PROGRAMA PARA LA VINCULACIÓN DE EMPRESAS CON INSTITUCIONES DE EDUCACION SUPERIOR Y CENTROS DE INVESTIGACIÓN COMECYT-EDOMEX CONVOCATORIA 2023" project number 34544 and Industrias MAR-ROB S.A. de C.V.

References

[1] Campero, M., Alarcón, L.F. (2018). *Administración de proyectos civiles*, Tercera edición, Ediciones UC, Valparaíso, Chile.

[2] Gido, J., Clements, J., Baker, R. (2003). *Administración exitosa de proyectos*, Cengage Learning Editores, Santa Fe, México, from *https://dspace.scz.ucb.edu.bo/dspace/bitstream/123456789/13442/3/9947.pdf,* accessed April 26, 2023.

[3] Turner, J.R. (2009). *Handbook of project-based management: Leading strategic change in organizations,* 3rd edition, Mc Graw Hill, New York, USA.

[4] Wan, J. (2022). Demand prediction and optimization of workshop manufacturing resources allocation: A new method and a case study, *Advances in Production Engineering & Management,* Vol. 17, No. 4, 413-424, doi: 10.14743/apem2022.4.445.

[5] Amaya Amaya, J. (2016). *Toma de decisiones gerenciales: Métodos cuantitativos para la administración,* 2nd edition, Ecoe Ediciones, Bogotá, Colombia, from Link, accessed April 26, 2023.

[6] Heizer, J.H., Render, B. (2004). *Principles of Operations Management*, 5th edition, Pearson Educación, Ciudad de México, Mexico, from Link, accessed April 26, 2023.

[7] Kerzner, H. (2013). *Project management: A systems approach to planning, scheduling, and controlling,* John Wiley, New York, USA, from Link, accessed April 27, 2023.

[8] Goodpasture, J.C. (2004). *Quantitative methods in project management,* J. Ross Publishing, Fort Lauderdale, Florida, USA, from Link, accessed April 27, 2023,.

[9] Saunders, M., Lewis, P. (2017). *Doing research in business and management*, 2nd edition, Pearson, Harlow, United Kingdom, from *https://research.birmingham.ac.uk/en/publications/doing-research-in-business-and-management,* accessed April 27, 2023.

[10] Kramar, D., Cica, D. (2021). Modeling and optimization of finish diamond turning of spherical surfaces based on response surface methodology and cuckoo search algorithm, *Advances in Production Engineering & Management,* Vol. 16, No. 3, 326-334, doi: 10.14743/apem2021.3.403.

[11] Turk, M., Pipan, M., Simic, M., Herakovic, N. (2020). Simulation-based time evaluation of basic manual assembly tasks, *Advances in Production Engineering & Management,* Vol. 15, No. 3, 331-344, doi: 10.14743/apem2020.3.369.

[12] Kan, S.H. (2004). *Metrics and models in software quality engineering* , 2nd edition, Pearson Education, New York USA, from Link, accessed April 27, 2023.

[13] Bryde, D., Broquetas, M., Volm, J.M. (2013). The project benefits of Building Information Modelling (BIM), *International Journal of Project Management*, Vol. 31, No. 7, 971-980, doi: 10.1016/J.IJPROMAN.2012.12.001.

[14] Ofori, G. (2010). Indicators for measuring construction industry development in developing countries, *Building Research & Information,* Vol. 29, No. 1, 40-50, doi: 10.1080/09613210150208787.

[15] Ward, Y., Graves, A. (2007). Through-life management: The provision of total customer solutions in the aerospace industry, *International Journal of Services, Technology and Management*, Vol. 8, No. 6, 455-477, doi: 10.1504/IJSTM.2007.013942.

[16] Burgess, T.F., McKee, D., Kidd, C. (2005). Configuration management in the aerospace industry: A review of industry practice, *International Journal of Operations and Production Management*, Vol. 25, No. 3, 290-301, doi: 10.1108/01443570510581880.

[17] Martínez Sánchez, A., Pérez Pérez, M. (2005), Supply chain flexibility and firm performance: A conceptual model and empirical study in the automotive industry, *International Journal of Operations & Production Management,* Vol. 25, No. 7, 681-700, doi: 10.1108/01443570510605090.

[18] Han, X., Zhao, P.X., Kong, D.X. (2022). A bi-objective optimization of airport ferry vehicle scheduling based on heuristic algorithm: A real data case study, *Advances in Production Engineering & Management,* Vol. 17, No. 2, 183-192, doi: 10.14743/apem2022.2.429.

[19] Vukelic, D., Simunovic, K., Kanovic, Z., Saric, T., Doroslovacki, K., Prica, M., Simunovic, G. (2022). Modelling surface roughness in finish turning as a function of cutting tool geometry using the response surface method, Gaussian process regression and decision tree regression, *Advances in Production Engineering & Management,* Vol. 17, No. 3, 367-380, doi: 10.14743/apem2022.3.442.

[20] Blair, M., Hartong, A. (2012). Multidisciplinary design tools for affordability, In: *Proceedings of 41st Structures, Structural Dynamics, and Materials Conference and Exhibit,* Atlanta, USA, 350-360, doi: 10.2514/6.2000-1378.

[21] Qian, L., Ben-Arieh, D. (2008). Parametric cost estimation based on activity-based costing: A case study for design and development of rotational parts, *International Journal of Production Economics,* Vol. 113, No. 2, 805-818, doi: 10.1016/J.IJPE.2007.08.010.

[22] Wang, H.S., Wang, Y.N., Wang, Y.C. (2013). Cost estimation of plastic injection molding parts through integration of PSO and BP neural network, *Expert Systems with Applications,* Vol. 40, No. 2, 418-428, doi: 10.1016/J.ESWA.2012.01.166.

[23] Verlinden, B., Duflou, J.R., Collin, P., Cattrysse, D. (2008). Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study, *International Journal of Production Economics,* Vol. 111, No. 2, 484-492, doi: 10.1016/J.IJPE.2007.02.004.

[24] Lallemand, H. (1999). Méthodes de chiffrage et indicateurs de la fonction devis, *Travail et méthodes,* No. 545, 35-44.

[25] Madan, J., Rao, P.V.M., Kundra, T.K. (2007). System for early cost estimation of die-cast parts, *International Journal of Production Research,* Vol. 45, No. 20, 4823-4847, doi: 10.1080/00207540600789016.

[26] Sung, B.S., Kim, I.S. (2008). The molding analysis of automobile parts using the die-casting system, *Journal of Materials Processing Technology,* Vol. 201, No. 1-3, 635-639, doi: 10.1016/J.JMATPROTEC.2007.11.185.

[27] Srivastava, A., Joshi, V., Shivpuri, R. (2004). Computer modeling and prediction of thermal fatigue cracking in die-casting tooling, *Wear*, Vol. 256, No. 1-2, 38-43, doi: 10.1016/S0043-1648(03)00281-3.

[28] Tsoukalas, V.D. (2008). Optimization of porosity formation in $AlSi_9Cu_3$ pressure die castings using genetic algorithm analysis, *Materials & Design,* Vol. 29, No. 10, 2027-2033, doi: 10.1016/J.MATDES.2008.04.016.

[29] Kim, D.-H., Kim, H.-G., Kim, H.-S. (2015). Design optimization and manufacture of hybrid glass/carbon fiber reinforced composite bumper beam for automobile vehicle, *Composite Structures,* Vol. 131, 742-752, doi: 10.1016/J.COPSTRUCT.2015.06.028.

[30] Singh, R., Sahay, A., Karry, K.M., Muzzio, F., Ierapetritou, M., Ramachandran, R. (2014). Implementation of an advanced hybrid MPC–PID control system using PAT tools into a direct compaction continuous pharmaceutical tablet manufacturing pilot plant, *International Journal of Pharmaceutics,* Vol. 473, No. 1-2, 38-54, doi: 10.1016/J.IJPHARM.2014.06.045.

[31] Liou, F., Slattery, K., Kinsella, M., Newkirk, J., Chou, H., Landers, R. (2007). Applications of a hybrid manufacturing process for fabrication of metallic structures, *Rapid Prototyping Journal,* Vol. 13, No. 4, 236-244, doi: 10.1108/13552540710776188.

[32] Rojek, I., Jagodziński, M. (2012). Hybrid artificial intelligence system in constraint based scheduling of integrated manufacturing ERP systems, In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.B. (eds.), In: *Hybrid artificial intelligent systems, HAIS 2012, Lecture notes in computer science,* Vol. 7209, Springer, Berlin, Germany, 229-240, doi: 10.1007/978-3-642-28931-6_22.

[33] Liu, N., Zhang, Y.F., Lu, W.F. (2015). A hybrid approach to energy consumption modelling based on cutting power: A milling case, *Journal of Cleaner Production*, Vol. 104, 264-272, doi: 10.1016/J.JCLEPRO.2015.05.049.

[34] Hao, Q., Shen, W. (2008). Implementing a hybrid simulation model for a Kanban-based material handling system, *Robotics and Computer-Integrated Manufacturing,* Vol. 24, No. 5, 635-646, doi: 10.1016/J.RCIM.2007.09.012.

[35] Langmaak, S., Wiseall, S., Bru, C., Adkins, R., Scanlan, J., Sóbester, A. (2013). An activity-based-parametric hybrid cost model to estimate the unit cost of a novel gas turbine component, *International Journal of Production Economics,* Vol. 142, No. 1, 74-88, doi: 10.1016/J.IJPE.2012.09.020.

[36] Zhu, Z., Dhokia, V.G., Nassehi, A., Newman, S.T. (2013). A review of hybrid manufacturing processes – state of the art and future perspectives, *International Journal of Computer Integrated Manufacturing,* Vol. 26, No. 7, 596-615, doi: 10.1080/0951192X.2012.749530.

[37] Shehab, E.M., Abdalla, H.S. (2001). Manufacturing cost modelling for concurrent product development, *Robotics and Computer-Integrated Manufacturing,* Vol. 17, No. 4, 341-353, doi: 10.1016/S0736-5845(01)00009-6.

[38] Shooter, S.B., Simpson, T.W., Kumara, S.R.T., Stone, R.B., Terpenny, J.P. (2005). Toward a multi-agent information management infrastructure for product family planning and mass customisation, *International Journal of Mass Customisation*, Vol. 1, No. 1, 134-155, doi: 10.1504/IJMASSC.2005.007354.

[39] Montgomery, D.C., Jennings, C.L., Kulahci, M. (2008). *Introduction to time series analysis and forecasting,* 2nd edition, Jonh Wiley & Sons, Hoboken, New Jersey, USA, doi: 10.1111/jtsa.12203.

[40] Rosato, D.V., Rosato, D.V., Rosato, M.G. (2012). *Injection molding handbook,* 3rd edition, Kluwer Academic Publishers, Norwell, Massachusetts, USA, doi: 10.1007/978-1-4615-4597-2.

APEM journal

Advances in Production Engineering & Management
Volume 18 | Number 2 | June 2023 | pp 175–186
https://doi.org/10.14743/apem2023.2.465

ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

When core sorting and quality grading is beneficial to remanufacturers: Insights from analytical models

Cao, X.H.[a], **Shi, X.L.**[a], **Lan, H.J.**[a], **Huang, D.**[a,*]

[a]School of Economics and Management, Beijing Jiaotong University, Beijing, P.R. China

A B S T R A C T

In this paper, we study the core acquisition and remanufacturing problem in which the remanufactured products are produced from acquired cores with uncertain quality condition, and are used to satisfy customer demand. Decision-making models are developed to examine the potential value of core sorting and quality grading in the remanufacturing system: a single-period model with deterministic demand, and a single-period model with stochastic demand (i.e., a newsvendor-type model). In each model, both the sorting strategy and the non-sorting strategy are discussed and compared. Our theoretical and numerical results show that: (1) In the deterministic demand case, core sorting is cost-effective only when the unit sorting cost is below a threshold value and the unit acquisition cost falls into a specific interval. Furthermore, in the case with two quality grades the adoption of sorting strategy with respect to the expected fraction of high-quality cores may be non-monotone: an initial increase in the expected fraction of high-quality cores may motivate a switch to core sorting, however, further increase in the expected fraction may motivate a reverse switch; (2) Similarly, in the stochastic demand case, the sorting strategy also becomes unattractive when the unit sorting cost is sufficiently high. In addition, the value of core sorting will be better off under more fluctuating demand for remanufactured products if the sorting strategy is the dominant strategy. Otherwise, it will be worse off.

A R T I C L E   I N F O

Keywords:
Remanufacturing;
Product (core) acquisition management;
Core sorting;
Quality grading;
Optimization;
Analytical models;
Cost-effectiveness

**Corresponding author:*
huangd@bjtu.edu.cn
(Huang, D.)

Article history:
Received 10 April 2023
Revised 12 June 2023
Accepted 17 June 2023

1. Introduction

Remanufacturing, which refers to the industrial process of restoring used products or parts to like-new condition, has become to be a more and more important part of the circular economy (Mihai *et al*., 2018), [1]. It also has been applied to various products such as automotive parts, electrical and electronics products, machinery, information and communication equipment, ink and toner cartridges, medical devices and furniture (Yoon and Joung, 2019; Davidavičienė *et al*., 2019, Yoon and Joung 2021), [2-4]. Many companies, including Daimler, Volvo, Lenovo, Huawei, Hewlett-Packard, Xerox, and IKEA, have engaged in the business of acquiring and remanufacturing used products (referred to as cores) from end-users. For example, the Volvo Group has reported that the total sales of remanufactured components amounted to SEK 10 billion in 2018 with a yearly average increasing rate of 10 %. Through an old-for-new program, Huawei has collected more than 140,000 used cell phones in 2018, and recovered and reused 82.3 % materials of these recycled cell phones. In a "Buy-back service" project launched in 2017, IKEA Japan recycled 1,900 second-hand furniture products in the first six months, and sold 85 % of them to customers after repair and refurbishing.

Since the quality condition of acquired cores can be so highly variable that may complicate the remanufacturing process (Guide, 2000; He, 2018), [5, 6], cores are usually sorted and graded into different categories based on their quality conditions by remanufacturers. For example, ReCellular, a cellphone remanufacturer, classified the returned phones into six quality grades based on their functional and physical criteria in which the remanufacturing cost of the lowest quality grade is eight times more than that of the highest quality grade (Guide *et al*., 2003), [7]. Similarly, a returned Personal Computer may also fall into different conditions: some may be brand new with their package box has never been opened; some may be used for a few times; some may need repair; and some may only be salvaged for parts or materials (Blackburn *et al*., 2004), [8]. It was reported that Pitney-Bowes, an equipment remanufacturer, categorized the returned mechanical products into three quality grades (i.e., remanufacturing, reassembling or recycling) according to different recovery options (Ferguson *et al*., 2009), [9]. In the reverse logistics network studied by Sedehzadeh and Seifbarghy (2021), [10], returned products are classified into two categories according to their health security: usable products and recoverable products. The sorting and grading operation resolves the quality uncertainties of acquired cores in advance, and thus allows remanufacturing these cores in a greedy and economical sequence.

In practice, core sorting and quality grading are typically implemented through visual inspection by sophisticated workers and usage information monitoring by information systems with tracking and tracing technology such as RFID. As the tasks of sorting and grading are primarily labour-intensive, it is usually costly to make such quality assessment. Furthermore, this prevailing operation helps firms to mitigate uncertainty in quality condition of cores, but requires acquiring more returned products than the demand for remanufactured products. As a result, the additional acquisition cost as well as the sorting cost offsets the remanufacturing cost savings obtained by processing a higher quality unit rather than a lower quality one. The trade-off between these two effects of core sorting and quality grading makes it less attractive in remanufacturing. Managers in remanufacturing firms are aware of the remanufacturing cost difference between different quality grades, but are unsure of the exact value of core quality information and the addition effort required to collecting and utilizing such information (Ferguson *et al*., 2009), [9]. Some researchers also found that under certain conditions remanufacturers can hardly benefit from core sorting and quality grading (see, e.g., Li *et al*., 2016; Yanıkoğlu and Denizel, 2021), [11, 12].

The research on core acquisition management, which refers to the management of the core acquisition process by dealing with the uncertainties with respect to return timing, quantity and core quality in reverse logistics, has intensified in recent years. An excellent review can be found in Wei *et al*. (2015), [13]. Our paper is most relevant to the research work that addressed the complications of quality uncertainty of cores in remanufacturing, thus we mainly review the literature of this research stream below. Guide *et al*. (2003), [7], first developed a single-period model to determine the optimal core acquisition prices and remanufactured products' selling price. They assumed that the collectors grade cores and sell them to the remanufacturer in different quality classes at different prices, and within a certain quality class all the cores have the same associated remanufacturing cost. Galbreth and Blackburn (2006), [14], considered the case where the remanufacturer acquires unsorted used products from the third party collectors. The quality condition of cores is highly variable, and the remanufacturer must sort each core into different quality grades. They assumed that the cumulative distribution of cores condition is known exactly. In a single-period model, they derived optimal core acquisition and sorting policies for both deterministic demand case and stochastic demand case. Their sorting policy is defined by the value of remanufacturing cost: cores with remanufacturing cost above a threshold value are scrapped, and those with cost below the threshold value are remanufactured. After that, Galbreth and Blackburn (2010), [15], studied core acquisition decisions when the quality condition of each acquired core is uncertain. Their work is extended to the case where the quality condition of each acquired core follows general distributions by Yang *et al*. (2015), [16]. By incorporating quantity discount and carbon tax scheme into the core acquisition models discussed in Galbreth and Blackburn (2006, 2010), [14, 15], Yang *et al*. (2016), [17], analysed the optimal core acquisition policies under the cases with quality variability and with both quality variability and condition uncertainty, respectively. Teunter and Flapper (2011), [18], assumed that there are multiple types of acquired cores, and that the type of a core follows a multinomial distribution. They derived optimal core acquisition and remanufacturing policies for both deterministic and stochastic demand. Mutha *et al*. (2016), [19], developed a two-period model to study core acquisition decisions when used products can be acquired either in bulk with uncertain quality levels, or in sorted grades with known quality levels, and the remanufacturer can acquire remanufacture cores before or after demand is realized. Lv *et al*. (2017), [20], discussed the two-period and multi-period manufacturing/remanufacturing models where the collection rate of used products can be influenced by a fixed investment. Mircea *et al*. (2023), [21] studied a repeated game in remanufacturing where used products are collected by the online/offline recycler and the manufacturer, and derived the optimal recycling prices for each collector. In a newsvendor-type model setting, Li *et al*. (2016), [11], analyzed the optimal decisions on core acquisition and remanufacturing quantities under both the remanufacturing-to-stock (RMTS) mode and the remanufacturing-to-order (RMTO) mode. They found that sorting may never be adopted in the RMTS mode regardless of the sorting cost. Ferguson *et al*. (2009), [9], and Yanıkoğlu and Denizel (2021), [12], studied the value of quality grading in the multi-period setting. The numerical results in Ferguson *et al*. (2009), [9], indicated that the ratio of return rates to demand rates, the number of quality grades, the distribution of the quality of returns and the cost difference between quality grades are the main drivers of the value of quality grading. Based on Ferguson *et al*. (2009)'s work [9], Yanıkoğlu and Denizel (2021), [12], study the core acquisition and remanufacturing problem under the case where the unit remanufacturing cost and unit resource requirement are uncertain. They found that when considering the unit grading cost, it may cause a significant deterioration in the value of grading, even make the grading totally useless.

This work is motivated by academic research (see, e.g., Ferrer and Ketzenberg, 2004, [22]; Ketzenberg *et al*., 2006, [23]; Ferguson *et al*., 2009, [9]; Li *et al*., 2016, [11]), and industrial practice (see, e.g., ReCellular, Pitney-Bowes, PneuLaurent, Caterpillar). We study the core acquisition and remanufacturing problem under a single-period setting in which the remanufactured products used to satisfy customer demand are produced from acquired cores with uncertain quality. Both the deterministic demand case and the stochastic demand case are discussed to examine the potential value of core sorting and quality grading in remanufacturing. In each case, two strategies are discussed and compared: (1) the sorting strategy, i.e., sorting and remanufacturing cores in a greedy sequence, and (2) the non-sorting strategy, i.e., remanufacturing cores in the natural sequence. During these different model settings, we aim to address the following questions: (1) Is core sorting and quality grading a cost-effective operation for remanufacturing firms? (2) Under what conditions does the remanufacturer benefit from core sorting and quality grading? (3) What are the impacts of demand and cost parameters on the value of core sorting and quality grading? (4) Whether these impacts change under different core acquisition and remanufacturing decision-making environments?

More specifically, we start with a single period model with deterministic demand, for which we derive the optimal core acquisition quantity for both the sorting and non-sorting strategy. By comparing with these two strategies, we find that core sorting and quality grading is cost-effective only when the unit sorting cost is sufficiently small and the unit core acquisition cost falls into a specific interval. Further analysis in a special case with two quality grades show that, an initial increase in the expected fraction of high-quality cores may motivate a switch to core sorting and quality grading, however, further increase may motivate a reverse switch. We then continue our study in a newsvendor-type model and characterize the optimal acquisition, sorting and remanufacturing policies. Similar to the deterministic demand case, core sorting and quality grading is beneficial to the remanufacturer only when the sorting cost is sufficiently low. Another observation is that when the sorting strategy is superior to the non-sorting strategy, it is more valuable under a higher level of demand variation.

The rest of this paper is organized as follows. In Section 2, we describe the basic problem setting and introduce the notations and assumptions used throughout the paper. In Section 3 and Section 4, we present a single-period core acquisition and remanufacturing model with deterministic and stochastic demand, respectively. Finally, we conclude the paper in Section 5.

2. Problem description

Consider a remanufacturing firm who acquires cores and sells remanufactured products to the market. The cores are acquired in bulk at unit acquisition cost c_r and their quality conditions are uncertain. Without loss of generality, assume that the acquired cores can be categorized into N nominal quality grades based on their quality levels, with grade 1 being the highest quality grade and grade N being the lowest. In general, the higher quality grades are cheaper to remanufacture, that is, $c_1 < c_2 <\ldots< c_N$, where c_i denotes the unit remanufacturing cost of grade i, $i = 1,2,\ldots,N$. The fraction of acquired cores belonging to grade i, θ_i, is a random variable with the domain [0, 1]. The expectation of θ_i is denoted by μ_i. The fractions $\theta_1, \theta_2,\ldots, \theta_N$ are jointly distributed with p.d.f. $g_{\theta_1,\theta_2,\ldots,\theta_N}(\cdot)$ and c.d.f. $G_{\theta_1,\theta_2,\ldots,\theta_N}(\cdot)$ and their summation is 1.

After a quantity of cores, Q units, are acquired, the firm has two alternative choices: (1) adopting a sorting procedure to resolve the quality uncertainties by incurring a unit sorting cost c_g, (2) doing nothing. If cores are sorted and graded, the fractions of core quality distribution are revealed before remanufacturing, otherwise they are known to the firm only until all the cores are remanufactured.

The market demand for remanufactured products studied in the single-period setting has two alternative forms: (a) deterministic demand, (b) stochastic demand. For the deterministic case, market demand D is known and no shortage is allowed. For the stochastic case, market demand D is a random variable with probability density function $f(D)$ and cumulative distribution function $F(D)$. In addition, we assume that D and $(\theta_1, \theta_2, \ldots, \theta_N)$ are independent. At the beginning of each period, the remanufacturing firm decides remanufacturing quantities from each grade of core before demand is realized. Then demand realization is observed. A unit shortage penalty cost, b_r, is charged for the unsatisfied demand and a unit holding cost, h_r, is charged for the leftover remanufactured product. At the end of the period, the leftover cores are disposed at unit cost, h_u. In general, the sequence of events is as follows: acquiring bulk cores, sorting and grading (or doing nothing), remanufacturing, demand realization. The decisions to be made by the remanufacturing firm include the core acquisition quantity and the remanufacturing quantity of each quality grade. The objective is to maximize the total expected profit with respect to core acquisition, remanufacturing, and demand fulfilment.

In summary, the following notions and assumptions are employed in this paper:

N	number of nominal quality grades of cores
D	market demand for remanufactured products
Q	acquisition quantity of cores
p	unit selling price for a remanufactured product
c_r	unit acquisition cost
c_i	unit remanufacturing cost of grade i, $i = 1,2,\ldots,N$
c_g	unit sorting cost
h_u	unit disposal cost
b_r	unit shortage penalty cost for unsatisfied demand
h_r	unit holding cost for the remanufactured product
θ_i	fraction of acquired cores belonging to grade i, $i = 1,2,\ldots,N$
μ_i	expected value of θ_i, $i = 1,2,\ldots,N$
c_μ	expected unit remanufacturing cost, where $c_\mu = E\left(\sum_{i=1}^{N} \theta_i c_i\right) = \sum_{i=1}^{N} \mu_i c_i$
$g_{\theta_1,\theta_2,\ldots,\theta_N}(\cdot)$	joint probability density function of $\theta_1, \theta_2, \ldots, \theta_N$
$G_{\theta_1,\theta_2,\ldots,\theta_N}(\cdot)$	joint cumulative distribution function of $\theta_1, \theta_2, \ldots, \theta_N$
$f(D)$	probability density function of D

$F(D)$	cumulative distribution function of D
q_i	quantity of cores of grade i that are remanufactured, $i = 1,2,\ldots,N$
π	total expected profit of the remanufacturing firm

3. The single-period, deterministic demand case

We begin our analysis with the single-period, deterministic demand case. Both the sorting scenario and the non-sorting scenario are discussed. For ease of exposition, we put superscript D and subscripts s and ns on some notations to represent these two scenarios, respectively.

3.1 The non-sorting scenario

When the manufacturer adopts the non-sorting operation, it is optimal to acquire the exact demanded quantity of cores. Let $Q_{ns}^{D^*}$ and $\pi_{ns}^{D^*}$ be the optimal acquisition quantity and the optimal expected profit under the non-sorting scenario, respectively. Then we have $Q_{ns}^{D^*} = D$ and $\pi_{ns}^{D^*} = (p - c_r - c_\mu)D$, where $c_\mu = \sum_{i=1}^{N} \mu_i c_i$ is the expected unit remanufacturing cost.

3.2 The sorting scenario

Let Q_s^D be the acquisition quantity of cores under the sorting scenario in the deterministic demand case. We define the acquisition ratio as $\alpha = D/Q_s^D$, where α can be interpreted as the fraction of acquired cores that must be remanufactured to satisfy the demand. It is reasonable that the remanufacturer will acquire more cores than demanded under the sorting scenario (Galbreth and Blackburn, 2006), [14]. Thus, we have $0 < \alpha \leq 1$.

The optimal remanufacturing policy is straightforward that follows a greedy rule: first processing cores of grade 1, and if the remanufacture-up-to level is still not reached, then processing cores of grade 2, and so on, stop until the remanufacture-up-to level is reached or all the cores are remanufactured.

Let $R_0 = 0$ and

$$R_i = min\left\{D, \sum_{j=1}^{i} \theta_j Q_s^D\right\}, i = 1,2,\ldots,N \tag{1}$$

Then the quantity of cores of grade i remanufactured under the sorting scenario, q_i^D, is

$$q_i^D = R_i - R_{i-1}, i = 1,2,\ldots,N \tag{2}$$

The expected profit of the remanufacturer, π_s^D, is

$$\pi_s^D = p \times D - (c_r + c_g)Q_s^D - E\left(\sum_{i=1}^{N} c_i q_i^D\right) - h_u(Q_s^D - D) \tag{3}$$

which can be rewritten as a function of α, where

$$\begin{aligned} \pi_s^D(\alpha) = {} & (p - c_N + h_u) \times D - (c_r + c_g + h_u)\frac{D}{\alpha} \\ & + D \times \sum_{i=1}^{N-1} (c_{i+1} - c_i)\left(1 - \int_0^\alpha (1 - \tfrac{x}{\alpha}) g_{\sum_{j=1}^{i} \theta_j}(x)dx\right) \end{aligned} \tag{4}$$

where $g_{\sum_{j=1}^{i}\theta_j}(x)$ and $G_{\sum_{j=1}^{i}\theta_j}(x)$ are the p.d.f. and c.d.f. of x, where $x = \sum_{j=1}^{i} \theta_j$ for any $i \in \{1,2,\ldots,N\}$.

Theorem 1. *If $c_r + c_g + h_u \geq c_N - c_\mu$, $\pi_s^D(\alpha)$ is monotonically increasing in α, and $\alpha^* = 1$. Otherwise, $\pi_s^D(\alpha)$ is quasi-concave in α, and α^* is given by*

$$\sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)\int_0^{\alpha^*} x g_{\sum_{j=1}^{i}\theta_j}(x)dx\right] = c_r + c_g + h_u \tag{5}$$

Proof: Take the first derivate of $\pi_s^D(\alpha)$ with respect to α, we have

$$\pi_s'^D(\alpha) = (c_r + c_g + h_u)\frac{D}{\alpha^2} - \frac{D}{\alpha^2}\sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)\int_0^{\alpha} xg_{\sum_{j=1}^{i}\theta_j}(x)dx\right]$$

It is easy to verify that the term $\sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)\int_0^{\alpha} xg_{\sum_{j=1}^{i}\theta_j}(x)dx\right]$ is increasing in α, and $\pi_s'^D(\alpha)$ is decreasing in α. Note that

(i) when $\alpha \to 0$, $\sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)\int_0^{\alpha} xg_{\sum_{j=1}^{i}\theta_j}(x)dx\right] \to 0$,

(ii) when $\alpha = 1$, $\sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)\int_0^{\alpha} xg_{\sum_{j=1}^{i}\theta_j}(x)dx\right] = \sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)\sum_{j=1}^{i}\mu_j\right] = c_N - c_\mu$.

If $c_r + c_g + h_u \geq c_N - c_\mu$, then we always have $\pi_s'^D(\alpha) \geq 0$ for any $\alpha \in (0,1]$. This implies that $\pi_s^D(\alpha)$ is increasing in α, and attains its maximum at $\alpha = 1$. If $c_r + c_g + h_u < c_N - c_\mu$, with the increase of α, $\pi_s'^D(\alpha)$ decreases and changes sign at most once from positive to negative. Denote α^* as the changing point where $\pi_s'^D(\alpha^*) = 0$. Thus, $\pi_s'^D(\alpha) > 0$ for any $\alpha \in (0, \alpha^*)$ and $\pi_s'^D(\alpha) < 0$ for any $\alpha \in (\alpha^*, 1]$. This implies that $\pi_s^D(\alpha)$ is increasing in α on the interval $(0, \alpha^*)$ and decreasing in α on the interval $(\alpha^*, 1]$. Therefore, $\pi_s^D(\alpha)$ is quasi-concave in α, and attains its maximum at $\alpha = \alpha^*$.□

Theorem 1 suggests that when the unit core processing (i.e., acquisition and sorting and disposal) cost is sufficiently high, the acquired cores are fully remanufactured. Otherwise, they are partially remanufactured. However, it is never optimal to adopt the sorting operation if all the acquired cores are remanufactured (i.e., $\alpha^* = 1$).

3.3 Comparative analysis

We examine the value of core sorting by comparing the expected profit of the remanufacturer under both the sorting scenario and the non-sorting scenario. Proposition 1 characterizes the situations when core sorting is beneficial to the remanufacturer.

Proposition 1. *There exist c_r^{max}, c_r^{min} and c_g^{max} such that (i) $\pi_s^D(\alpha^*) > \pi_{ns}^{D^*}$, if $c_r^{min} < c_r < c_r^{max}$ and $c_g < c_g^{max}$; (ii) $\pi_s^D(\alpha^*) \leq \pi_{ns}^{D^*}$, otherwise.*

Proof. By Proposition 1, when $c_r + c_g + h_u \geq c_N - c_\mu$, $\alpha^* = 1$, i.e., all the acquired cores are remanufactured. Then it is never optimal for the remanufacturer to choose the sorting strategy since additional sorting cost is required but no remanufacturing cost savings are obtained from core sorting. Thus, under this case we have $\pi_s^D(\alpha^*) < \pi_{ns}^{D^*}$.

When $c_r + c_g + h_u < c_N - c_\mu$,

$$\pi_s^D(\alpha^*) - \pi_{ns}^{D*} = (c_r + c_\mu - c_1 + h_u)D - D\sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)G_{\sum_{j=1}^{i}\theta_j}(\alpha^*)\right]$$

Thus, if $c_r > c_1 + \sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)G_{\sum_{j=1}^{i}\theta_j}(\alpha^*)\right] - c_\mu - h_u$, then $\pi_s^D(\alpha^*) > \pi_{ns}^{D*}$; otherwise, $\pi_s^D(\alpha^*) \leq \pi_{ns}^{D^*}$.

Let $c_r^{max} = c_N - c_\mu - c_g - h_u$, and $c_r^{min} = c_1 + \sum_{i=1}^{N-1}\left[(c_{i+1} - c_i)G_{\sum_{j=1}^{i}\theta_j}(\alpha^*)\right] - c_\mu - h_u$. Note that under this case the necessary condition for $\pi_s^D(\alpha^*) > \pi_{ns}^{D^*}$ is $c_r^{max} > c_r^{min}$, i.e., $c_g < \sum_{i=1}^{N-1}\left\{(c_{i+1} - c_i)[1 - G_{\sum_{j=1}^{i}\theta_j}(\alpha^*)]\right\} = c_g^{max}$.

Proposition 1 implies that core sorting in remanufacturing is cost-effective only when the unit sorting cost is smaller than a threshold and the unit core acquisition cost falls into a specific interval.

Furthermore, we examine the two-grade case to gain more managerial insights. It is found that the adoption of core sorting with respect to the expected fraction of high-quality cores, μ_1, may be non-monotone: an initial increase in μ_1 may motivate a switch to core sorting, however, further increase in μ_1 may motivate a reverse switch.

Proposition 2. *In the case with two quality grades, there exist μ_1^{max}, μ_1^{min} and c_g^{max} such that (i) $\pi_s^D(\alpha^*) > \pi_{ns}^{D^*}$, if $\mu_1^{min} < \mu_1 < \mu_1^{max}$ and $c_g < c_g^{max}$; (ii) $\pi_s^D(\alpha^*) \leq \pi_{ns}^{D^*}$, otherwise.*

Proof. According to the proof of Proposition 1, when $N = 2$, from $c_r^{min} < c_r < c_r^{max}$ we have

$$\mu_1^{min} = \frac{c_r + c_g + h_u}{c_2 - c_1} < \mu_1 < 1 - G_{\theta_1}(\alpha^*) + \frac{c_r + h_u}{c_2 - c_1} = \mu_1^{max}$$

Furthermore, from $\mu_1^{max} > \mu_1^{min}$ we have $c_g < (c_2 - c_1)[1 - G_{\theta_1}(\alpha^*)] = c_g^{max}$.

The results of Proposition 2 are illustrated in Fig. 1, where the parameter settings in the associated numerical example are: $D = 50$, $p = 100$, $c_r = 2$, $c_g = 2$, $h_u = 1$, $c_1 = 5$, $c_2 = 30$. The fraction of acquired cores with high-quality level, θ_1, follows a Beta distribution, i.e., $\theta_1 \sim \mathrm{B}(a, b)$. We consider different values of a and b in the numerical study, where $(a, b) \in \{a = 0.5i, b = 10 - 0.5i, i = 1, \ldots, 19\}$, thus the expectation of θ_1, μ_1, changes its value from 0.05 to 0.95 in steps of 0.05. The impacts of the expected fraction of high-quality cores, μ_1, on the expected profit of the remanufacturer are shown in Fig. 1. The explanation is as follows. When the fraction of high-quality cores is excessively high, there is no need to sorting cores as most of the acquired cores are in good quality condition. While when the fraction of high-quality cores is excessively low, core sorting could not help to improve the poor quality condition of acquired cores and thus is useless.

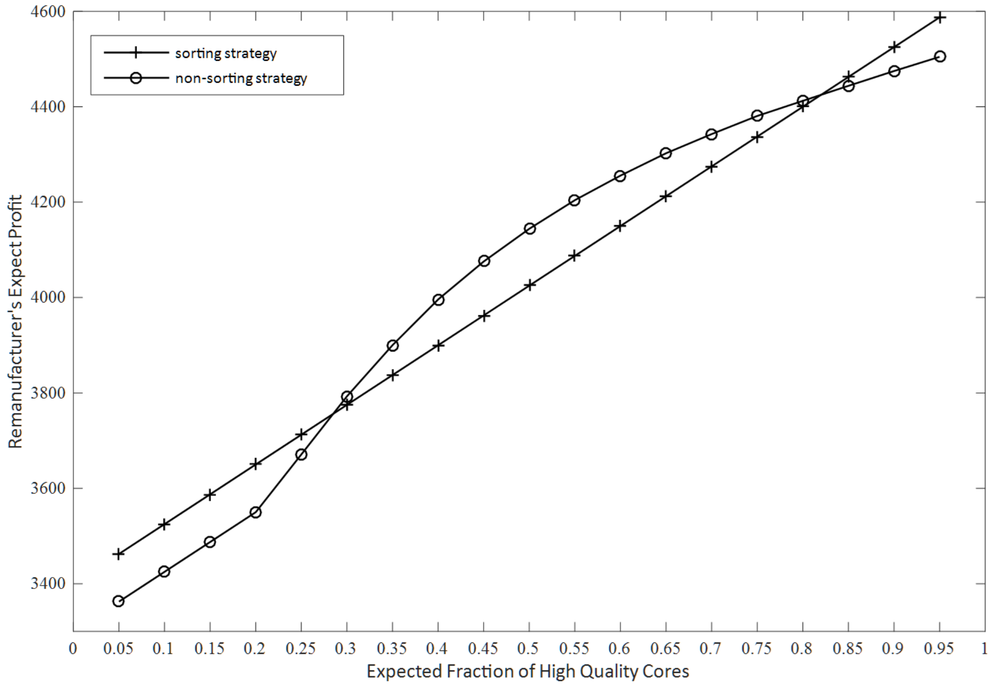


Fig. 1 Remanufacturer's expected profit under sorting and non-sorting strategies at different levels of expected fraction of high-quality cores

4. The single-period, stochastic demand case

Then we extend our analysis to the single-period, stochastic demand case. Similarly, we use superscript R and subscripts s and ns to represent this case under sorting and non-sorting scenarios.

4.1 The non-sorting scenario

In the non-sorting scenario, without the quality information of cores in advance, the firm will process all the acquired cores to satisfy market demand. The decision-making problem is a newsvendor-type problem. The expected profit of the firm is

$$\pi_{ns}^R = E[pmin(Q_{ns}^R, D) - b_r(D - Q_{ns}^R)^+ - h_r(Q_{ns}^R - D)^+] - (c_\mu + c_r)Q_{ns}^R \tag{6}$$

The optimal acquisition quantity, $Q_{ns}^{R^*}$, is

$$Q_{ns}^{R^*} = F^{-1}\left(1 - \frac{c_r + c_\mu + h_r}{p + b_r + h_r}\right) \tag{7}$$

4.2 The sorting scenario

Let Q_s^R be the acquisition quantity of the cores, and q_i^R the quantity of cores of grade i remanufactured. The expected profit of the firm under the sorting scenario, is

$$\pi_s^R = E[pmin(Q_s^R, D) - \sum_{i=1}^{N} c_i q_i^R - b_r(D - \sum_{i=1}^{N} q_i^R)^+ - h_r(\sum_{i=1}^{N} q_i^R - D)^+ - h_u(Q_s^R - \sum_{i=1}^{N} q_i^R)] \quad (8)$$
$$-(c_r + c_g)Q_s^R$$
$$s.t. \quad 0 \le q_i^R \le \theta_i Q_s^R, i = 1,2,\ldots,N$$

Similarly, let $R_0 = 0$ and R_i be the aggregate remanufactured quantity from grade 1 to grade i cores, $i = 1,2,\ldots,N$. Then following the greedy rule, $q_i^R = R_i - R_{i-1}$. The expected profit (8) can be rewritten as

$$\pi_s^R = E[pmin(Q_s^R, D) - \sum_{i=1}^{N} c_i(R_i - R_{i-1}) - b_r(D - R_N)^+ - h_r(R_N - D)^+ - h_u(Q_s^R - R_N)] \quad (9)$$
$$-(c_r + c_g)Q_s^R$$
$$s.t. \quad 0 \le R_i \le \textstyle\sum_{k=1}^{i} \theta_k Q_s^R, i = 1,2,\ldots,N$$

The remanufacture-up-to levels to the unconstrained optimization problem (9) are characterized in Proposition 3, which are also newsvendor-type solutions.

Proposition 3. *The remanufacture-up-to level from grade* 1 *to grade i core,* R_i^*, *is*

$$R_i^* = F^{-1}\left(1 - \frac{c_i + h_r - h_u}{p + b_r + h_r}\right), i = 1,2,\ldots,N \quad (10)$$

where $R_1^* > R_2^* > \ldots > R_N^*$.

Proof. If only cores of grade 1 to grade i ($i = 1,2,\ldots,N$) are remanufactured, then the remanufacturer's expected profit (9) can be rewritten as

$$\pi_s^R = E[pmin(R_i, D) - \sum_{l=1}^{i} c_l(R_l - R_{l-1}) - b_r(D - R_i)^+ - h_r(R_i - D)^+ - h_u(Q_s^R - R_i)] - (c_r + c_g)Q_s^R$$

Taking the first and second derivative of π_s^R with respect to R_i respectively, we have

$$\frac{\partial \pi_s^R}{\partial R_i} = (p + b_r + h_u - c_i) - (p + b_r + h_r)F(R_i)$$

and

$$\frac{\partial^2 \pi_s^R}{\partial {R_i}^2} = -(p + b_r + h_r)f(R_i) < 0$$

Thus, π_s^R is concave in R_i and the optimal remanufacture-up-to level is given by the first-order conditions. Moreover, since $c_1 < c_2 < \ldots < c_N$, it is easy to verify that $R_1^* > R_2^* > \ldots > R_N^*$.□

Let y_i^* be the optimal aggregate remanufactured quantity from grade 1 to grade i cores after remanufacturing, then the optimal remanufacturing quantity of cores of grade i is $q_i^{R^*} = y_i^* - y_{i-1}^*$, where $y_0^* = 0$. Based on the remanufacture-up-to levels derived in Proposition 3, the optimal remanufacturing policy under the sorting strategy with stochastic demand is characterized in Theorem 2.

Theorem 2. *After core sorting, the fractions* $\theta_1, \theta_2, \ldots, \theta_N$ *are realized, then for a given core acquisition quantity* Q_s^R, *the optimal remanufacturing policy,* $\{y_1^*, \ldots, y_i^*, \ldots, y_N^*\}$, *is as follows:*

(1) if $R_1^* \le \theta_1 Q_s^R$, $y_N^* = \ldots = y_1^* = R_1^*$;
(2) for $i = 2,\ldots,N$,
if $\sum_{k=1}^{i-1} \theta_k Q_s^R < R_i^* \le \sum_{k=1}^{i} \theta_k Q_s^R$, $y_N^* = \ldots = y_i^* = R_i^*$, $y_l^* = \sum_{k=1}^{l} \theta_k Q_s^R$, $l = 1,2,\ldots,i-1$;
if $R_i^* \le \sum_{k=1}^{i-1} \theta_k Q_s^R < R_{i-1}^*$, $y_N^* = \ldots = y_i^* = y_{i-1}^*$, $y_l^* = \sum_{k=1}^{l} \theta_k Q_s^R$, $l = 1,2,\ldots,i-1$;
(3) if $Q_s^R < R_N^*$, $y_i^* = \sum_{k=1}^{i} \theta_k Q_s^R$, $i = 1,2,\ldots,N$.

Proof. According to Proposition 3, when cores of grade 1 to grade i ($i = 1,2,\dots,N$) are remanufactured, the optimal remanufacture-up-to level is R_i^*, which is given by Eq. (10). However, for a given core acquisition quantity Q_S^R and the realized fractions $\theta_1, \theta_2, \dots, \theta_N$, the available quantity of grade 1 to grade i cores is $\sum_{k=1}^{i} \theta_k Q_S^R$. Therefore, the optimal aggregate remanufactured quantity from grade 1 to grade i cores after remanufacturing, y_i^*, depends on the relationship between R_i^* and $\sum_{k=1}^{i} \theta_k Q_S^R$.

When $R_i^* > \sum_{k=1}^{i} \theta_k Q_S^R$, we have $R_1^* > \dots > R_{i-1}^* > R_i^* > \sum_{k=1}^{i} \theta_k Q_S^R > \sum_{k=1}^{i-1} \theta_k Q_S^R > \dots > \theta_1 Q_S^R$, then it is optimal to make grade 1 to grade i cores fully remanufactured. Otherwise, it is optimal to remanufacture up to R_i^*, however, if $R_i^* < \sum_{k=1}^{i-1} \theta_k Q_S^R$, then grade i cores will never be remanufactured.

Theorem 2 implies that in the optimal remanufacturing policy there exists a certain quality grade so that the acquired cores from grades that are lower than this grade are fully remanufactured, and the acquired cores from grades that are higher than this grade are disposed of, while the acquired cores of this grade are partially remanufactured.

Some properties of π_S^R are derived in Proposition 4, which can facilitate us to derive the optimal core acquisition quantity.

Proposition 4. $\pi_S^R(Q_S^R)$ *has the following properties*:
(i) $\pi_S^R(Q_S^R)$ *is piece-wise concave and continuous with respect to* Q_S^R;
(ii) $\pi_S^R(Q_S^R)$ *has a unique global optimum*.

Proof. The proof follows along the lines of proof in Brown and Lee (2003), [24]. In Eq. (8), since $\{Q_S^R : Q_S^R \geq 0\}$ is a convex set, $\{q_1^R, \dots, q_N^R : 0 \leq q_i^R \leq \theta_i Q_S^R, i = 1,2,\dots,N\}$ is a nonempty set, and $-\pi_S^R$ is a convex function on the nonempty convex set, then following Proposition A.4 in Porteus (2002), [25], $\min_{0 \leq q_i^R \leq \theta_i Q_S^R} -\pi_S^R(Q_S^R)$ is convex in Q_S^R, i.e., $\max_{0 \leq q_i^R \leq \theta_i Q_S^R} \pi_S^R(Q_S^R)$ is concave in Q_S^R. Hence, $\pi_S^R(Q_S^R)$ has a unique global optimum. Moreover, based on Theorem 2 it is easy to show that $\pi_S^R(Q_S^R)$ is also a piece-wise continuous function with respect to Q_S^R.□

Fig. 2 illustrates the results of Proposition 4. The parameter settings of the numerical example are: $N = 2$, $D \sim N(100,20)$, $p = 100$, $c_r = 2$, $c_g = 2$, $h_r = 2$, $b_r = 5$, $h_u = 1$, $c_1 = 6$, $c_2 = 30$, $\theta_1 \sim B(8,2)$. It can be seen from Fig. 2 that $\pi_S^R(Q_S^R)$ is divided into three pieces by $Q_S^R = R_1^*$ and $Q_S^R = R_2^*$, and each piece is a concave function.

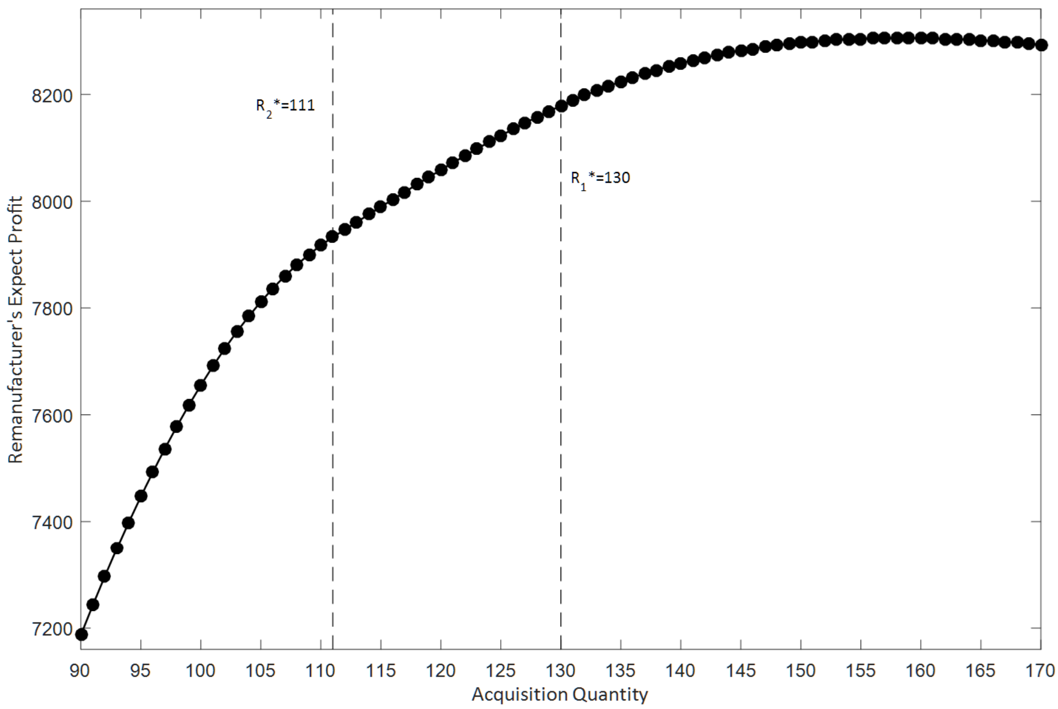


Fig. 2 Remanufacturer's expected profit with different core acquisition quantities

4.3 Comparative analysis

We also examine the value of core sorting under the stochastic demand case. By comparison, it is also found that core sorting in remanufacturing is cost-effective only when the unit sorting cost is below a threshold value and the unit core processing (i.e., acquisition, sorting and disposal) cost is sufficiently low.

Proposition 5. *There exists* c_g^{max} *such that* (i) $\pi_s^R(Q_s^{R^*}) > \pi_{ns}^R(Q_{ns}^{R^*})$, *if* $c_r + c_g + h_u < c_N - c_\mu$ *and* $c_g < c_g^{max}$; (ii) $\pi_s^R(Q_s^{R^*}) < \pi_{ns}^R(Q_{ns}^{R^*})$, *otherwise.*

Proof. It follows Theorem 2 that when $Q_s^R \leq R_N^*$, $y_i^* = \sum_{k=1}^{i} \theta_k Q_s^R$, $i = 1,2,\ldots,N$. Then, on the interval $[0, R_N^*]$ we have

$$\pi_s^R = E[pmin(Q_s^R, D) - b_r(D - Q_s^R)^+ - h_r(Q_s^R - D)^+] - (c_\mu + c_r + c_g)Q_s^R$$

Taking the first derivative of π_s^R with respect to Q_s^R, we have

$$\pi_s'^R = (p + b_r - c_r - c_g - c_\mu) - (p + b_r + h_r)F(Q_s^R)$$

It is easy to verify that $\pi_s'^R$ is decreasing in Q_s^R. Moreover, when $Q_s^R = 0, \pi_s'^R(Q_s^R = 0) > 0$; when $Q_s^R = R_N^*, \pi_s'^R(Q_s^R = R_N^*) = c_N - c_\mu - (c_r + c_g + h_u)$.

When $c_r + c_g + h_u \geq c_N - c_\mu$, $\pi_s'^R(Q_s^R = R_N^*) \leq 0$, thus the optimal core acquisition quantity $Q_s^{R^*}$ is on the interval $[0, R_N^*]$. This implies that all the acquired cores should be remanufactured even with the sorting strategy. As mentioned above, under this condition core sorting has no value to the remanufacturer.

When $c_r + c_g + h_u < c_N - c_\mu$, $\pi_s'^R(Q_s^R = R_N^*) > 0$, thus π_s^R is increasing in Q_s^R on the interval $[0, R_N^*]$ since $\pi_s'^R$ is always positive on that interval. This suggests that $Q_s^{R^*}$ is larger than R_N^*, i.e., at least the acquired cores of grade N are partially remanufactured. Note that if $c_g = 0$, then for any given quantity of cores, the sorting strategy is not inferior to the non-sorting strategy due to the potential remanufacturing cost savings. Thus, we have $\pi_s^R(Q_s^{R^*}) > \pi_s^R(Q_{ns}^{R^*}) \geq \pi_{ns}^R(Q_{ns}^{R^*})$, where the first inequality is valid due to the optimality of $Q_s^{R^*}$ for π_s^R. With the increase of c_g, $\pi_s^R(Q_s^{R^*})$ decreases but $\pi_{ns}^R(Q_{ns}^{R^*})$ remains the same. Therefore, there exists a value of c_g on the interval$(0, c_N - c_\mu - c_r - h_u)$, denoted as c_g^{max}, so that when$0 < c_g < c_g^{max}$, $\pi_s^R(Q_s^{R^*}) > \pi_{ns}^R(Q_{ns}^{R^*})$.□

Proposition 5 implies that when the sorting cost is too high, core sorting is useless to the remanufacturer. It also can be seen from Table 1 that when $c_g \leq 2$, the remanufacturer's expected profit under the sorting strategy is higher, while when $c_g \geq 2.5$, the non-sorting strategy makes the remanufacturer more profitable.

We further examine the impacts of demand uncertainty on the value of core sorting. Let Δ_g be the increase in remanufacturer's expected profit by using the sorting strategy relative to the case where the remanufacturer adopts the non-sorting strategy, where $\Delta_g = \frac{\pi_g - \pi_m}{\pi_m} \times 100\%$. Based on the same demand and cost parameter settings with that of Fig. 2, we vary the values of a and b (distribution parameters of θ_1) where $(a, b) \in \{a = 0.5i, b = 10 - 0.5i, i = 1,\ldots,19\}$, and consider three different levels of demand variation: $\delta_D \in \{10, 20, 30\}$.

Fig. 3 illustrates the increase in remanufacturer's expected profit by adopting the sorting strategy under different levels of demand variation. The observation is similar to the deterministic demand case that the sorting strategy is valuable to remanufacturer only when the expected fraction of high-quality cores is within a suitable interval. Another observation is that when the sorting strategy is superior to the non-sorting strategy, it is more valuable under a higher level of demand variation. This is because a higher level of demand variation may increase the possibility that a high-quality core is remanufactured rather than being disposed of, which facilitates the adoption of sorting strategy.

Table 1 Remanufacturer's expected profit under sorting and non-sorting strategies at different levels of sorting cost

c_g	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
π_s^R	8633.3	8549.0	8466.4	8385.2	8305.4	8226.7	8149.1	8072.5	7996.8	7921.9	7847.9
π_{ns}^R	8238.8	8238.8	8238.8	8238.8	8238.8	8238.8	8238.8	8238.8	8238.8	8238.8	8238.8

Note: Other parameters take the same value as in the numerical example of Fig. 2.

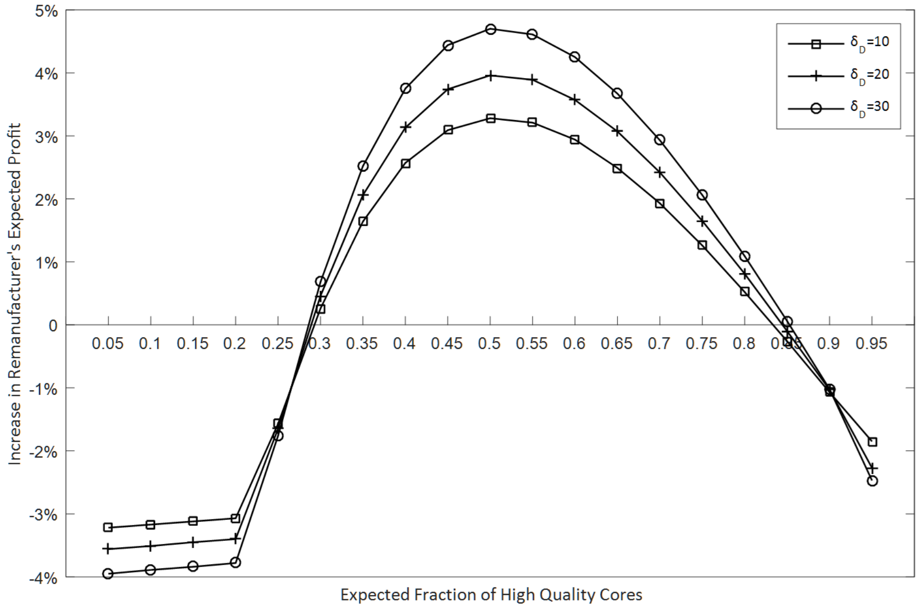


Fig. 3 Increase in Remanufacturer's expected profit under sorting and non-sorting strategies at different levels of demand variation

5. Conclusion

In this paper, we study the core acquisition and remanufacturing problem under a single-period setting in which the remanufactured products are produced from acquired cores with uncertain quality. Both the deterministic demand case and the stochastic demand case are discussed to examine the potential value of core sorting and quality grading in remanufacturing. For deterministic demand, it is found that core sorting is cost-effective only when the sorting cost is below a threshold value and the acquisition cost falls into a specific interval. Furthermore, in the case with two quality grades the sorting strategy is valuable to remanufacturer only when the expected fraction of high-quality cores is within a suitable interval. For stochastic demand, it is observed that the sorting strategy also becomes unattractive when the sorting cost is sufficiently high. In addition, the value of core sorting will be larger with more variable demand for remanufactured products if the sorting strategy is the dominant strategy. Future research may include considering the core acquisition and remanufacturing problem in a multi-period setting, and/or incorporating acquisition pricing decisions into the models.

Acknowledgement

The authors would like to thank the editor and anonymous reviewers for their valuable comments and suggestions, which help to improve the quality of this paper greatly.

References

[1] Mihai, M., Manea, D., Titan, E., Vasile, V. (2018). Correlations in the European circular economy, *Economic Computation and Economic Cybernetics Studies and Research,* Vol. 52, No. 4, 61-78, doi: 10.24818/18423264/52.4.18.05.

[2] Yoon, J., Joung, S. (2019). Examining purhcase intention of eco-friendly products: A comparative study, *Journal of System and Management Sciences*, Vol. 9, No. 3, 123-135, doi: 10.33168/JSMS.2019.0308.

[3] Davidavičienė, V., Raudeliūnienė, J., Zubrii, M. (2019). Evaluation of customers' sustainable fashion perception, *Journal of System and Management Sciences*, Vol. 9, No. 4, 50-66, doi: 10.33168/JSMS.2019.0405.

[4] Yoon, J., Joung, S. (2021). Environmental self-identity and purchasing eco-friendly products, *Journal of Logistics, Informatics and Service Sciences*, Vol. 8, No. 1, 82-99, doi: 10.33168/LISS.2021.0106.

[5] Guide, V.D.R. (2000). Production planning and control for remanufacturing: Industry practice and research needs, *Journal of Operations Management*, Vol. 18, No. 4, 467-483, doi: 10.1016/S0272-6963(00)00034-6.

[6] He, P. (2018) Optimization and simulation of remanufacturing production scheduling under uncertainties, *International Journal of Simulation Modelling*, Vol. 17, No. 4, 734-743, doi: 10.2507/IJSIMM17(4)CO20.

[7] Guide, V.D.R, Teunter, R.H., van Wassenhove, L.N. (2003). Matching demand and supply to maximize profits from remanufacturing, *Manufacturing & Service Operations Management*, Vol. 5, No. 4, 303-316, doi: 10.1287/msom.5.4.303.24883.

[8] Blackburn, J.D., Guide, V.D.R, Souza, G.C., van Wassenhove, L.N. (2004). Reverse supply chains for commercial returns, *California Management Review*, Vol. 46, No. 2, 6-22, doi: 10.2307/41166207.

[9] Ferguson, M., Guide, V.D., Koca, E., Souza, G.C. (2009). The value of quality grading in remanufacturing, *Production and Operations Management*, Vol. 18, No. 3, 300-314, doi: 10.1111/j.1937-5956.2009.01033.x.

[10] Sedehzadeh, S., Seifbarghy, M. (2021). Redesigning a closed loop food supply chain network considering sustainability and food banks with different returns, *Economic Computation and Economic Cybernetics Studies and Research,* Vol. 55, No. 4, 85-100, doi: 10.24818/18423264/55.4.21.06.

[11] Li, X., Li, Y., Cai, X. (2016). On core sorting in RMTS and RMTO systems: A newsvendor framework, *Decision Sciences*, Vol. 47, No. 1, 60-93, doi: 10.1111/deci.12152.

[12] Yanıkoğlu, İ., Denizel, M. (2021) The value of quality grading in remanufacturing under quality level uncertainty, *International Journal of Production Research,* Vol. 59, No. 3, 839-859, doi: 10.1080/00207543.2020.1711983.

[13] Wei, S., Tang, O., Sundin, E. (2015). Core (product) acquisition management for remanufacturing: A review, *Journal of Remanufacturing*, Vol. 5, Article No. 4, doi: 10.1186/s13243-015-0014-7.

[14] Galbreth, M.R., Blackburn, J.D. (2006). Optimal acquisition and sorting policies for remanufacturing, *Production and Operations Management*, Vol. 15, No. 3, 384-392, doi: 10.1111/j.1937-5956.2006.tb00252.x.

[15] Galbreth, M.R., Blackburn, J.D. (2010). Optimal acquisition quantities in remanufacturing with condition uncertainty, *Production and Operations Management*, Vol. 19, No. 1, 61-69, doi: 10.1111/j.1937-5956.2009.01067.x.

[16] Yang, C.-H., Wang, J., Ji, P. (2015). Optimal acquisition policy in remanufacturing under general core quality distributions, *International Journal of Production Research*, Vol. 53, No. 5, 1425-1438, doi: 10.1080/00207543.2014.944283.

[17] Yang, C.-H., Bao, X.-Y., Song, C., Liu, H.-B. (2016). Optimal acquisition policy in remanufacturing systems with quantity discount and carbon tax scheme, *Tehnički Vjesnik – Technical Gazette,* Vol. 23, No. 4, 1073-1081, doi: 10.17559/TV-20160521140622.

[18] Teunter, R.H., Flapper, S.D.P. (2011). Optimal core acquisition and remanufacturing policies under uncertain core quality fractions, *European Journal of Operational Research*, Vol. 210, No. 2, 241-248, doi: 10.1016/j.ejor.2010.06.015.

[19] Mutha, A., Bansal, S., Guide, V.D.R. (2016). Managing demand uncertainty through core acquisition in remanufacturing, *Production and Operations Management*, Vol. 25, No. 8, 1449-1464, doi: 10.1111/poms.12554.

[20] Lv, X., Huang, J.-H., Liu, H.-B. (2017). Optimal manufacturing/remanufacturing policies with fixed investment for the underdeveloped remanufacturing system, *Tehnički Vjesnik – Technical Gazette*, Vol. 24, No. 5, 1491-1499, doi: 10.17559/TV-20170829170855.

[21] Mircea, G., Neamţu, M., Sîrghi, N., Ştefea, P. (2023) The dynamical analysis of the sustainability of a recycling mathematical model, *Economic Computation and Economic Cybernetics Studies and Research,* Vol. 57, No. 1, 41-56, doi: 10.24818/18423264/57.1.23.03.

[22] Ferrer, G., Ketzenberg, M.E. (2004). Value of information in remanufacturing complex products, *IIE Transactions*, Vol. 36, No. 3, 265-277, doi: 10.1080/07408170490274223.

[23] Ketzenberg, M.E., van der Laan, E., Teunter, R.H. (2006). Value of information in closed loop supply chains, *Production and Operations Management*, Vol. 15, No. 3, 393-406, doi: 10.1111/j.1937-5956.2006.tb00253.x.

[24] Brown, A.O., Lee, H.L. (2003). The impact of demand signal quality on optimal decisions in supply contracts, In: Shanthikumar, J.G., Yao, D.D., Zijm, W.H.M. (ed.), *Stochastic modeling and optimization of manufacturing systems and supply chains*, Springer, Boston, USA, 299-328, doi: 10.1007/978-1-4615-0373-6_12.

[25] Porteus, E. (2002). *Foundation of stochastic inventory theory*, Stanford University Press, Redwood City, California, USA, doi: 10.1515/9781503619883.

Advances in Production Engineering & Management **ISSN 1854-6250**
Volume 18 | Number 2 | June 2023 | pp 187–198 Journal home: apem-journal.org
https://doi.org/10.14743/apem2023.2.466 Original scientific paper

Engineering-to-order manufacturing: A criticality analysis of key challenges and solutions based on literature review

Fortes, C.S.[a,b,*], Tenera, A.B.[b,c], Cunha, P.F.[a], Teixeira, J.P.[b]

[a]Escola Superior de Tecnologia de Setúbal (ESTS), Instituto Politécnico de Setúbal, Portugal
[b]UNIDEMI, DEMI, NOVA School of Science & Technology (FCT), Universidade NOVA de Lisboa, Portugal
[c]LASI, Laboratório Associado de Sistemas Inteligentes, Portugal

A B S T R A C T

Engineer-to-Order (ETO) manufacturing companies involve customised production products based on specific customer requirements, and face a significant challenge. Some of those challenges are related with this type of company's activities in production scheduling, planning and control, efficiency improvement and lead time reduction. The present study was conducted with a systematic literature review and a survey from ETO firms to identify the most frequent and critical problems. Among the most critical issues identified is the difficulty in optimising production performance (P3), with a GCI value of 16, implying that both time and cost share the same critical level. An analysis using a proposed Criticality Matrix was then performed enabling companies to prioritise decision-making and resource allocation. The results highlight the importance of adopting mass customisation strategies, innovative approaches and workflow optimisation. Continuous monitoring and analysis of criticality levels can also help ETO companies identify emerging issues and improve informed decisions. Effective communication and collaboration among stakeholders were also identified as vital. Future research could be done expanding further the study sample and developing decision-support tools for ETO manufacturing companies. This study contributes to the field by providing a new criticality matrix for ETO companies to understand better and address their production challenges, aiding in decision-making and resource allocation.

A R T I C L E I N F O

Keywords:
Communication and collaboration;
Critical factors;
Decision-making;
Engineer-to-order;
Literature review;
Production planning and control;
Production scheduling;
Resource allocation;
Workflow optimisation

**Corresponding author:*
Carlos.fortes@estsetubal.ips.pt
(Fortes, C.S.)

Article history:
Received 5 June 2023
Revised 23 June 2023
Accepted 28 June 2023

1. Introduction

1.1 Background

Engineering-to-Order (ETO) companies typically manufacture highly complex and customised products with significant engineering components that provide high value. As highlighted by Jiang and Xi, these companies utilise an ETO strategy particularly when complex structures are required to be built [1]. In general, and particularly in European Union, such companies are small or medium-sized and play essential roles in specific sectors and regions, primarily operating in international markets. ETO companies' economic development and competitiveness are crucial for their success, as they supply various industrial sectors, such as the automotive, aeronautical, and pharmaceutical/medical device industries. Their business model is centred around their ability to design, develop and manufacture products to order, delivering value to their clients [2]. One example of a sector where ETO companies are common is the mould manufacturing industry, especially tools and plastics. ETO companies face many challenges in production planning and shop floor management as product and process diversity and complexity increase system variability. These highly competitive sectors require the efficient execution of numerous activities to meet customer expectations.

Therefore, shop floor planning and operational decision-making effectiveness are essential for these companies. In addition, they must fulfil increasingly shorter delivery times defined by the market or clients, who also demand quality and lower costs. ETO companies often must also deal with a workflow that requires the quick generation of product designs, detailed bill of material and manufacturing work instructions for each order, to meet tailored customer specifications.

1.2 Scope and outline

In the context of Engineer-to-Order (ETO), companies encounter diverse tasks, high-level customisation of products, and an unpredictable market. These elements call for insightful planning and decision-making, seamlessly integrated with essential operations such as tendering and procurement, to enhance performance across their value chain. Consequently, the impetus for this investigation is threefold: (a) to assist ETO companies in honing their efficiency and effectiveness in operations, augmenting their impact and importance in a specific industrial domain and the broader economy; (b) to introduce novel strategies, methodologies, procedures, or resources to ETO companies to boost their market competitiveness; and (c) to lessen the innate challenges associated with ETO operations, particularly by enhancing their capacity to manage the volatility and intricacy within their entire value chain. A SLR was conducted to identify the main challenges and issues related to ETO companies and the methods and approaches presented in the scientific literature to resolve or mitigate these issues in the value chain of this type of company. The search for scientific literature on the topic aimed to gain knowledge about various viewpoints and holistically interpret the study's theories and models. Additionally, ETO companies, were surveyed to identify the most frequent and critical problems, with a subsequent criticality analysis using a developed criticality matrix for problems in ETO companies, offering valuable insights into industry perspectives and potential avenues for improvement. This paper is structured as follows: Section 1 presents an overview of the study's background and motivation.

2. Materials and methods

2.1 Research methodology

An SLR was developed to identify ETO companies' difficulties and the technical and scientific solutions proposed to address them, followed by a survey of ETO companies to assess the criticality of the identified problems in those organisations.

Unlike narrative reviews, the SLR aim to answer a specific research question through a planned and structured approach to identifying, selecting, and critically appraising relevant studies [3]. By analysing and synthesising different authors findings, a comprehensive understanding of the existing body of work is gained, gaps to explore are identified, and conclusions are drawn on what is known and not known [4]. The review followed the PRISMA [5] guidance approach, with specific eligible criteria that included: papers discussing ETO problems, process planning, manufacturing related articles, production planning and control, and articles published in English. Articles that did not meet these criteria were excluded. The relevant studies were identified by searching digital databases, such as Web of Science and Scopus, as shown in Table 1. Articles published between January 2017 and April 2022 were the ones in this study considered, using search terms such as "engineer* to order" AND "variab*" OR "uncertaint*" OR "complex*" AND "produc* control" OR "plan*" OR "schedul*" OR "produc* process" OR "workflow*" OR "shop-floor*" OR "digital*" OR "problem*" OR "issue*" OR "constraint*" OR "inefficient*" NOT "ship*" OR "building*".

Table 1 Databases used for the search

Database	Search Terms
Web of Science	TS = ("engineer* to order") AND TS = (variab* OR uncertaint* OR complex*) AND TS = ("produc* control" OR plan* OR schedul* OR "produc* process" OR workflow* OR "shop-floor*" OR digital* OR problem* OR issue* OR constraint* OR inefficient*) NOT TS = (ship* OR building*)
Scopus	TITLE-ABS-KEY("engineer* to order") AND TITLE-ABS-KEY(variab* OR uncertaint* OR complex*) AND TITLE-ABS-KEY("produc* control" OR plan* OR schedul* OR "produc* process" OR workflow* OR "shop-floor*" OR digital* OR problem* OR issue* OR constraint* OR inefficient*) AND NOT(ship* OR building*) AND (LIMIT-TO(DOCTYPE,"ar") OR LIMIT-TO(DOCTYPE,"cp")) AND (LIMIT-TO(LANGUAGE,"English"))

The study search yielded 83 WoS publications and 57 Scopus publications in a total of 140 potentially relevant publications, from which 32 were duplicated as shown in Fig. 1. From the initial 108 publications analysed for inclusion, 62 articles did not meet the selection criteria due to their publication dates falling outside of the 2017-2022 range, thus leading to their exclusion. This process yielded 46 potentially most relevant references to undergo a comprehensive in-depth analysis, out of which only 18 were deemed pertinent. Further to this rigorous analysis, six articles were identified and subsequently included in the study. Fig. 1 shows the flow of citations through the systematic review process.

The final sample size of 24 papers can be considered a robust number for this analysis, as similar sample sizes have been used in other studies such as 19 papers in nursing studies [6], 23 papers in learning management [7], and 23 papers in project management studies [8]. In the following section, the results of the in-depth content analysis of the identified articles are synthesised, focusing on the general framework and competitive environment in which ETO operates and, in the challenges and issues faced by this type of company, including its principal characteristics and organisation.

In addition to the SLR, a survey was also conducted among ETO companies to better understand the frequency and the criticality of the identified problems. This survey aimed to check the findings of the SLR and provide any new real-world perspectives on the challenges encountered in ETO environments. The collected data were analysed using a proposed criticality matrix to assess the identified problems potentially severe impact and prioritise potential solutions. This approach allowed for a more comprehensive understanding of all the identified problems and issues faced by ETO companies, providing an informed potential action.

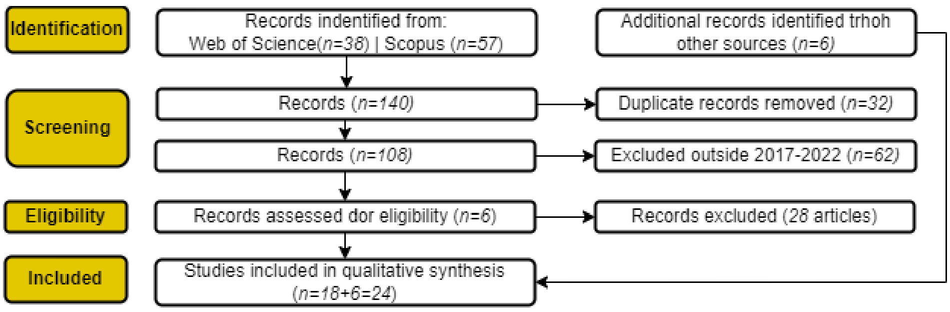


Fig. 1 A schematic flow chart diagram of the selected studies

2.2 A concept map approach of ETOs challenges and solutions

ETO companies specialise in fabricating high-value-added customised products in low volumes, typically small batches or one-of-a-kind products, as shown in Fig. 2.

As project-based organisations, ETO companies focus mainly on designing and manufacturing new products, involving a complex and highly specialised process that demands qualified and flexible workers [9]. Moreover, ETO companies are customer-centric [10], involving strong customer participation in product development. However, due to the initial phase's complexity and uncertainty in customer product characteristics, raw material purchase and product manufacturing only begin after the customer confirms the order quotation.

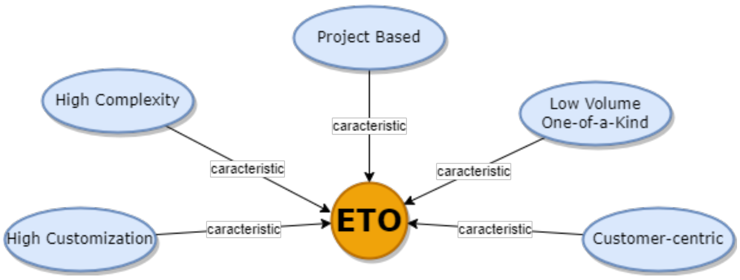


Fig. 2 Main attributes of ETO

ETO Potential key challenges

The demand for increasingly customised products has resulted in the need for better performance in ETO companies [11]. However, the discrete nature of shop floor workflows in ETO companies poses a significant management challenge requiring higher decision-making resilience to workflow performance [12]. The tactical-level management of value chain information can also be considered another critical issue in the dynamic environment of ETO companies [13]. During the contracting stage, uncertainties about costs and delivery date forecasts are typical, resulting in higher manufacturing costs [14], longer lead times, higher inventory costs, transportation costs, and other supply risks. Planning in ETO companies is complex, given the combination of multiple technical responsibility areas and the difficulty in predicting workload resources and lead times [15]. Customers can be involved in all project stages, such as design, production, assembly and testing [16]. These challenges often result in manufacturing process delays due to changes made during the project's progress [17]. Late changes in engineering design can cause costs increases and affect resource allocation along the value chain, according to Gosling *et al.* [18]. The involvement of customers and the various interactions place a high demand on production planning and control activities [19]. Also, as noted by Akcay *et al.* [20]. One-of-a-kind products or low-volume batches increase the number of jobs pending in manufacturing.

Developing a concept map for scientific and technological solutions

A thorough comprehension of ETO organisations and their challenges can be significantly enhanced by adopting innovative and engaging visual representations, such as concept maps. Unlike traditional tables, these maps can provide an intuitive and interactive means of conveying detailed information.

Derived from the problem table resulting from the SLR conducted in the research, an ETO domain, concept map can be proven beneficial for identifying and examining ETOs diverse issues, weaknesses, and areas warranting further investigation [21]. From Fortes *et al.* [21], Fig. 3, depicts a multi-layered concept map, with the ETO companies and their characteristics placed at the centre, encircled by a second ring that contains the surrounding issues. The third ring introduces the authors' proposals, and the outermost ring emphasises the contributions and limitations of the studies. This layout was design to enhance the understanding of problem areas, underline the need for further research, and to promote the development of adaptive strategies to tackle ETO organisations' multifaceted challenges. This concept map employs five distinct colours and three unique shapes to differentiate identified distinct categories. Blue balloons represent the characteristics of ETO companies; white balloons signify the issues to be addressed; yellow balloons indicate approaches; green balloons denote contributions; and red balloons symbolise weaknesses.

For example, the item "Bid solution uncertain", as presented in the study by Sylla *et al.* [22], is represented by a white balloon, signifying the issue that needs to be addressed. Subsequent yellow rectangles illustrate the approach of "a multi-criteria approach to bid solution uncertainty". For example, a green balloon showcases contributions, such as "providing bid solutions with accurate and timely replies", whilst a red balloon points out weaknesses, like "single case study with limited data". This visual approach intends to enable readers to effortlessly identify the categories of each item and their relationships with one another. Furthermore, connections between balloons can be established to illustrate the identified links between items. For instance, a connection between balloons one and two can be drawn, as both pertain to uncertain bid solutions.

So, concept maps offer a compelling and effective way of organising and presenting complex information [23] related to ETO organisations' challenges and potential solutions. By employing a visually appealing and interactive layout, these maps can facilitate a deeper understanding of the issues, approaches, contributions, and weaknesses within this field, ultimately encouraging further investigation and development of comprehensive strategies to address the dynamic challenges faced by ETO companies.

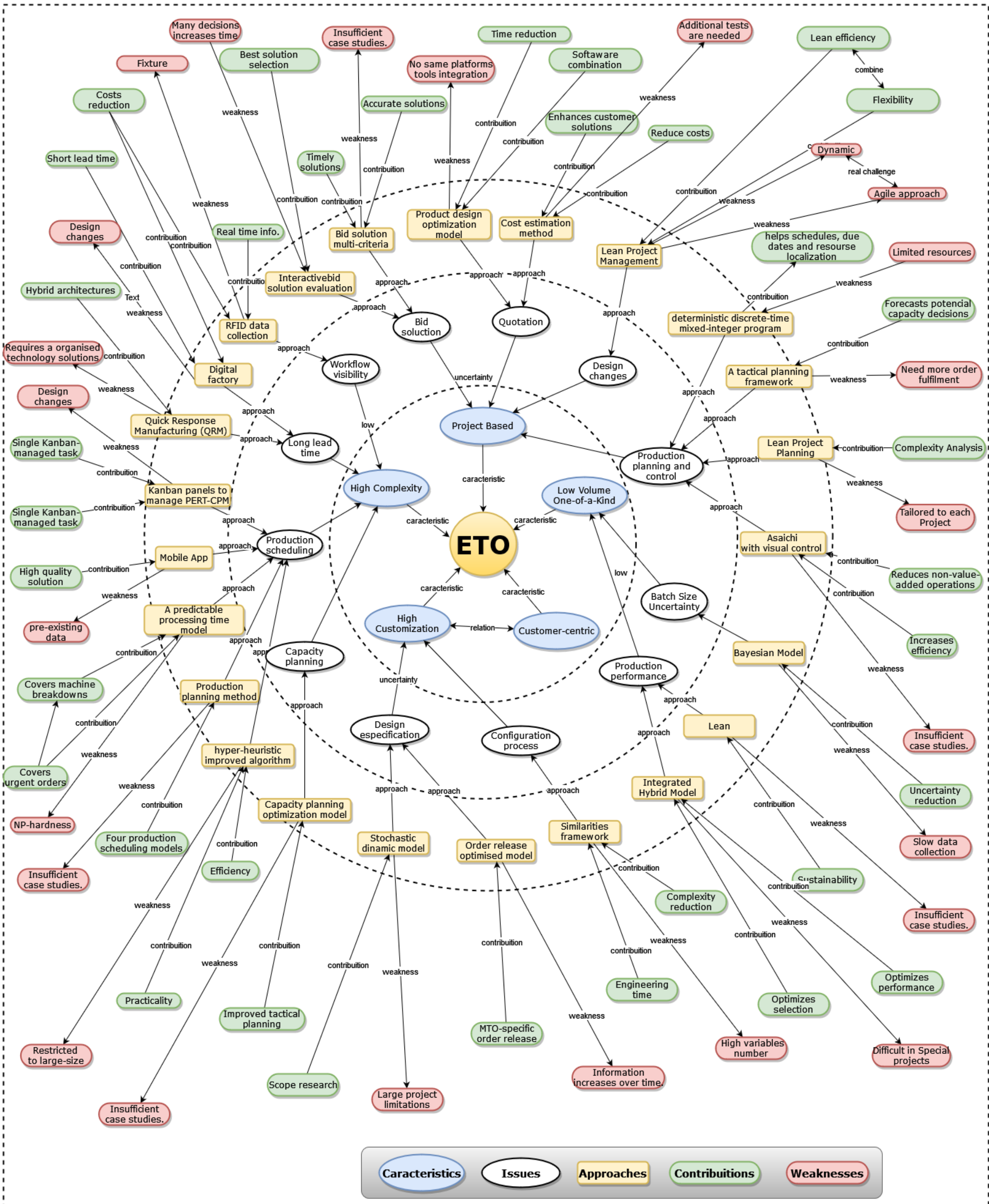


Fig. 1 ETO concept diagram [21]

3. Results and discussion: Criticality analysis of ETO companies' problems

The systematic literature review (SLR) results and ETOS's companies problem assessment survey examined the frequency and identified problems impact these organisations encounter, establishing the associated criticality. To ensure a diverse sample survey was developed specifically targeting high-level professionals such as managers, department directors, and project managers from a diverse range of ETO companies, including different industry sectors, sizes, and geographical locations. This investigation uncovered new valuables insights, enabling stakeholders to better understand the pressing concerns affecting ETO businesses. In addition, insights into pressing issues in ETO companies, allowing stakeholders to prioritise efforts, develop effective strategies, and improve performance and competitiveness within the sector can be provided by the proposed "Criticality Analysis of ETO Company Problems" table.

3.1 Problem frequency assessment

The analysis involved assessing the relative citation frequencies of each problem identified in the SLR *cf* and comparing them to the average organisational occurrence rankings reported by the study survey *fr* to check the theoretical known SLR findings and identify any new current problems. This study's results discovered no additional problems, and all problems were reported as occurring in practice. Table 2 lists the 14 problems discovered through the SLR, where problem P10 (Production Scheduling) emerged as the most cited, followed by P11 (Production Planning and Control), with P2 (Tender Proposal), P3 (Production Performance), P12 (Product Specification), and P14 (Lead Time) in third place. In terms of reported frequency *fr* all problems were assessed as existing validating the SLR identified problems and revealing that 12 of the 14 problems (P1, P2, P4, P5, P7, P8, P9, P10, P11, P12, P13, and P14) were equally frequently assessed with P3 (Production Performance) the most frequent problem and P6 (Unstructured Knowledge) the least concerned from the surveyed companies' perspective.

Table 2 Criticality analysis of ETO company problems

Item	Designation	Description	*pf* (%)	Ranking	*fr* (%)	Ranking	Deviation
P1	Capacity Planning	Difficulty in adapting existing capacities to new product development and production plans.	4	4th	7	2nd	+2
P2	Tender Proposal	Limitations in accurately defining tender proposals due to uncertainty.	8	3rd	7	2nd	+1
P3	Production Performance	Difficulty in process optimisation.	8	3rd	10	1st	+2
P4	Workflow Visibility	Difficulty in visualising workload and supporting decision-making in a dynamic shop floor.	4	4th	7	2nd	+2
P5	Configuration Process	Frequent constraints on workflow configuration due to resource scarcity.	4	4th	7	2nd	+2
P6	Unstructured Knowledge	Inefficient reuse of design knowledge for new products.	4	4th	5	3rd	+1
P7	Lot Size Uncertainty	Uncertainty of lot size.	4	4th	7	2nd	+2
P8	Quotation	Difficulty in product design optimisation, influencing quotations.	4	4th	7	2nd	+2
P9	Workflow	Information flow failures, limiting collaboration and operational performance.	4	4th	7	2nd	+2
P10	Production Scheduling	Difficulty in defining optimal job sequences for production.	21	1st	7	2nd	-1
P11	Production Planning & Control	Addressing Variability in Operation Execution Times in Manufacturing and Assembly Processes	17	2nd	7	2nd	0
P12	Product Specification	Lack of information and detail in customer-provided product specifications.	8	3rd	7	2nd	+1
P13	Customer Changes	Late design changes due to customer-imposed alterations.	4	4th	7	2nd	+2
P14	Lead Time	Long execution time.	8	3rd	7	2nd	+1

pf – published frequencies (Scopus, Web of Science); *fr* – frequency as reported in the survey; *Dif* – difference: *Dif* = *fp* – *fr*

So, upon examining the table, several conclusions can be drawn regarding the criticality of problems ETO companies face. First, comparing published frequencies *pf* from the Scopus and Web of Science databases with the frequencies reported in the survey *fr*, which included 12 participating ETOs companies, reveals some discrepancies between the literature and the real-world challenges experienced by these firms.

Table 2 results highlight that Production Scheduling (P10) and Production Planning & Control (P11) are the most critical issues for ETO companies, as evidenced by their high published frequencies (21 % and 17 %, respectively) and their shared second positions in the survey. These findings, indicate that these issues are consistently recognised as significant challenges in academic literature and industry practice. However, most of the problems display a positive deviation in their rankings between the published frequencies *pf* and the frequencies reported in the survey (*fr*), suggesting that the participating ETO companies perceive these issues to be more critical than what is reflected in the literature. This discrepancy could imply that the existing literature must fully capture ETO companies' unique challenges or that ETO practitioners are more aware of the issues within their sector. Considering these findings, further research is recommended to explore the specific challenges faced by ETO companies in greater depth and foster closer collaboration between practitioners and academics. This collaboration can help ensure that the literature accurately reflects the practical issues experienced in the field. By addressing this gap, ETO companies can better understand and mitigate their challenges, improving their overall performance and competitiveness.

Therefore, by focusing on the most frequent concerned issues, such as Production Scheduling (P10) and Production Planning & Control (P11), researchers can contribute to developing more practical solutions and strategies, enhancing the sector's ability to tackle the complex problems inherent in ETO operations.

Moreover, the detailed examination of the impact on the field of each problem can provide valuable insights into the prioritization of challenges faced by ETO companies, guiding future research and decision-making in the sector.

3.2 Analysis of ETO problem criticality

This session analysis the critical [24] level of the ETO companies' problems to prioritise ETOs organisations' challenges. Here, the criticality assessment approach is based on the probability measure by the frequency assessment and impact of the identified problems. It is employed to assist companies in better understanding the challenges they face and develop targeted strategies for addressing them. Furthermore, this type of analysis can allow a more effective allocation of resources and informed decision-making, improving the overall performance and success of ETO projects.

Table 3 presents the proposed matrix for evaluating the criticality level of the identified ETOs problems. The criticality index *CI* results from the intersection of the row probability *P* and the column impact *I*, where the values are multiplied, thus obtaining the criticality level through the general Eq. 1.

$$CI = P \times I \tag{1}$$

Table 3 Risk matrix (Probability/Impact matrix)

Probability (P)	5	0	5	10	15	20	25
	4	0	4	8	12	16	20
	3	0	3	6	9	12	15
	2	0	2	4	6	8	10
	1	0	1	2	3	4	5
	0	0	0	0	0	0	0
Probability/Impact		0	1	2	3	4	5
		Impact (*I*)					
Non-existent/Very Low [0, 4]	Low [4, 6]	Moderate [6, 12]	High [12, 16]		Very High [16, 25]		

The criticality level of problems may vary among organisations, resulting in different evaluations based on each company's specific situation, as demonstrated in Table 4.

Table 4 Specific organisational evaluation results

Issue ETO1	*P*	cI_{ETO1}	tI_{ETO1}	*OCI*	Issue ETO2	*P*	cI_{ETO2}	tI_{ETO2}	*OCI*	Issue ETO3	*P*	cI_{ETO3}	tI_{ETO3}	*OCI*
P1	2	2	4	8	P1	2	5	4	10	P1	2	4	4	8
P2	2	3	4	8	P2	2	4	2	8	P2	2	2	2	4
P3	3	3	5	15	P3	2	5	2	10	P3	4	5	2	20
P4	2	1	4	8	P4	2	2	3	6	P4	2	4	3	8
P5	3	4	4	12	P5	3	5	4	15	P5	4	5	4	20
P6	1	1	3	3	P6	2	1	2	4	P6	3	1	2	6
P7	1	1	2	2	P7	2	2	2	4	P7	2	4	2	8
P8	2	3	4	8	P8	2	5	5	10	P8	2	2	5	10
P9	1	4	5	5	P9	2	3	5	10	P9	2	4	5	10
P10	1	1	2	2	P10	2	3	1	6	P10	2	5	1	10
P11	2	3	3	6	P11	2	3	1	6	P11	2	5	1	10
P12	3	1	4	12	P12	3	5	5	15	P12	2	5	5	10
P13	3	4	5	15	P13	2	5	5	10	P13	2	5	5	10
P14	3	5	1	15	P14	3	5	1	15	P14	3	2	1	6
ETO4	*P*	cI_{ETO4}	tI_{ETO4}	*OCI*	ETO5	*P*	cI_{ETO5}	tI_{ETO5}	*OCI*	ETO6	*P*	cI_{ETO6}	tI_{ETO6}	*OCI*
P1	3	5	4	15	P1	3	3	3	9	P1	4	3	5	20
P2	3	3	3	9	P2	2	3	3	6	P2	3	2	4	12
P3	4	3	3	12	P3	3	2	3	9	P3	5	4	5	25
P4	3	4	3	12	P4	3	3	3	9	P4	5	3	5	25
P5	4	4	3	16	P5	3	3	2	9	P5	4	2	5	20
P6	4	3	1	12	P6	3	3	3	9	P6	2	1	2	4
P7	3	3	4	12	P7	3	3	3	9	P7	2	1	2	4
P8	4	4	3	16	P8	3	2	2	6	P8	4	2	2	8
P9	4	4	4	16	P9	3	3	3	9	P9	2	3	5	10
P10	4	2	4	16	P10	3	3	3	9	P10	2	2	3	6
P11	3	3	4	12	P11	4	3	2	12	P11	2	2	4	8
P12	4	4	4	16	P12	2	2	3	6	P12	5	2	5	25
P13	4	4	5	20	P13	2	3	2	6	P13	4	2	5	20
P14	3	4	5	15	P14	2	3	3	6	P14	5	3	4	20
ETO7	*P*	cI_{ETO7}	tI_{ETO7}	*OCI*	ETO8	*P*	cI_{ETO8}	tI_{ETO8}	*OCI*	ETO9	*P*	cI_{ETO9}	tI_{ETO9}	*OCI*
P1	3	3	3	9	P1	4	4	5	20	P1	2	2	4	8
P2	1	1	1	1	P2	4	4	5	20	P2	2	3	2	6
P3	5	4	3	20	P3	4	3	5	20	P3	3	4	5	15
P4	3	1	1	3	P4	5	5	5	25	P4	3	4	5	15
P5	1	2	2	2	P5	3	3	3	9	P5	2	4	4	8
P6	1	1	1	1	P6	1	1	1	1	P6	1	4	5	5
P7	1	1	2	2	P7	1	1	1	1	P7	5	3	4	20
P8	1	1	1	1	P8	4	5	4	20	P8	5	4	4	20
P9	3	1	1	3	P9	4	4	4	16	P9	5	4	5	25
P10	2	2	2	4	P10	1	1	1	1	P10	3	3	3	9
P11	3	3	2	9	P11	3	3	5	15	P11	2	3	5	10
P12	2	3	1	6	P12	3	3	4	12	P12	2	2	2	4
P13	2	2	4	8	P13	3	3	4	12	P13	3	4	4	12
P14	1	1	1	1	P14	4	3	5	20	P14	3	3	4	12
ETO10	*P*	cI_{ET10}	tI_{ET10}	*OCI*	ETO11	*P*	cI_{ET11}	tI_{ET11}	*OCI*	ETO12	*P*	cI_{ET12}	tI_{ET12}	*OCI*
P1	5	4	5	25	P1	4	4	4	16	P1	2	4	4	8
P2	4	3	4	16	P2	4	5	4	20	P2	3	5	4	15
P3	5	5	5	25	P3	3	4	5	15	P3	3	5	4	15
P4	5	5	4	25	P4	4	3	3	12	P4	3	5	4	15
P5	5	5	5	25	P5	3	4	5	15	P5	3	5	4	15
P6	2	1	1	2	P6	3	2	2	6	P6	3	5	4	15
P7	4	1	1	4	P7	4	4	3	16	P7	2	4	4	8
P8	2	2	4	8	P8	4	4	4	16	P8	3	5	4	15
P9	2	2	2	4	P9	4	4	4	16	P9	3	5	4	15
P10	3	4	4	12	P10	4	5	4	20	P10	3	5	4	15
P11	5	5	5	25	P11	5	5	5	25	P11	3	5	4	15
P12	2	2	2	4	P12	4	5	4	20	P12	2	5	4	10
P13	4	3	2	12	P13	4	5	4	20	P13	3	5	4	15
P14	4	4	4	16	P14	4	5	5	20	P14	3	5	4	15

P – Probability; *cI* – Cost Impact; *tI* – Time Impact; *OCI* – Organizational Criticality Index

Therefore, their specific values may differ from the general average trend values. The Organisational Criticality Index *OCI* results from the multiplication of the problem reported frequency with the highest value among its reported impacts on costs cI(x) or time tI(x). For the analysis of the general criticality index *GCI* of each identified problem, the approximate average level of reported occurrence frequency *of* was similarly multiplied by the average impacts on costs cI and time tI for each identified problem using Eqs. 2 to 4.

$$of = \frac{(frETO1 + \ ... + \ fr\ ETOn)}{n} \tag{2}$$

$$cI = \frac{(cIETO1 + \ .... + cIETOn)}{n} \tag{3}$$

$$tI = \frac{(tIETO1 + \ ... \ + \ tIETOn)}{n} \tag{4}$$

Table 5 summarises the criticality levels obtained for each problem, identifying the impact variable that most contributed to its value through time t or cost c.

Based on the findings from Table 5, the following conclusions can be drawn in the context of ETO companies. First, our analysis of the *GCI* highlights the key challenges that significantly impact these organisations in terms of time t and cost c. This understanding can enable ETO companies to develop targeted strategies and address these critical issues effectively, thereby enhancing their overall performance and competitiveness in the market.

These results demonstrate that several problems substantially impact ETO companies' performance. As noted by Leksic, the performance of an ETO company is directly tied to the efficiency of its production process, reinforcing the magnitude of the identified problems in this study [25]. Among the most critical issues identified is the difficulty in optimising production performance (P3), with a *GCI* value of 16*tc* meaning that both time and cost share the same critical level. This challenge arises from the highly customised nature of ETO products, which necessitates constant adaptation of manufacturing processes to meet the unique requirements of each project. Consequently, achieving the desired quality standards may require increased time and cost.

Capacity planning (P1) is another significant challenge that the surveyed ETOs companies face, with a *GCI* value of 12*tc*. This issue stems from the difficulty in aligning existing capacities with planning new product development and their introduction into production. As a result, ETO companies must diligently allocate resources and manage production schedules to cater to each customer's specific needs, which may pressure the overall capacity planning process.

Moreover, the analysis emphasises the importance of the configuration process (P5), which has a *GCI* value of 12*tc*. Frequent constraints in the workflow configuration process may occur due to the scarcity of resources concerning planned and ongoing works. The customer-centric approach of ETO companies exacerbates this issue, as accommodating unique customer requirements may lead to frequent changes and adjustments in the workflow configuration.

Table 5 Level of General Criticality Index (*GCI*)

12 ETO	Average Frequency Index of ETO Survey			*GCI*
Item	*fo*	*cI*	*tI*	
P1	3	4	4	*12tc*
P2	3	3	3	*9tc*
P3	4	4	4	*16tc*
P4	3	3	4	*12t*
P5	3	4	4	*12tc*
P6	2	2	2	*4tc*
P7	3	2	3	*9t*
P8	3	3	4	*12t*
P9	3	3	4	*12t*
P10	3	3	3	*9tc*
P11	3	4	3	*12c*
P12	3	3	4	*12t*
P13	3	4	4	*12tc*
P14	3	4	3	*12c*

tI - time impact; cI - cost impact

Additionally, the impact of customer-imposed changes on product design (P13) features prominently in our findings, with a *GCI* value of 12*tc*. Late changes in product design can disrupt the manufacturing process, resulting in delays and increased costs. Therefore, ETO companies must establish strategies to manage and minimise the impact of these changes on their operations.
Our analysis results also indicate that several other challenges possess moderate to high levels of criticality for ETO companies. These include workflow visibility (P4), the uncertainty of batch size (P7), pricing (P8), information flow failures (P9), production scheduling (P10), production planning and control (P11), product specification (P12), and lead time (P14). Each of these challenges is directly or indirectly influenced by the unique characteristics of ETO companies, such as their customer-centric approach, project-based operations, high customisation, low volume, extended lead times, high complexity, and high variability.

In conclusion, our examination of the criticality index for various issues faced by ETO companies emphasises the need for these organisations to recognise and address the challenges arising from their distinctive operational characteristics. Furthermore, by comprehending the criticality of these issues concerning time and cost, ETO companies can devise targeted strategies and solutions to improve their overall performance [26].

Potential strategies to tackle these challenges may include adopting lean management practices, enhanced communication and coordination among different departments, investment in employee training and development, and implementing advanced tools and technologies to boost process efficiency [27]. By addressing the critical issues identified in our study and considering the unique characteristics of ETO companies, it is feasible to enhance operational efficiency, reduce costs and lead times, and ultimately increase customer satisfaction and market competitiveness.

4. Conclusion

Analysing ETO companies' challenges highlights the importance of understanding and addressing critical factors that influence ETO manufacturing systems. By adopting mass customisation strategies and assessing problem criticality, ETO companies might more efficiently allocate resources, thereby improving time and cost management. Furthermore, focus on resource management, innovative approaches, and workflow optimisation will allow ETO companies to adapt to customer demands, increasing satisfaction and competitiveness. This study conducted an SLR and survey of ETO companies to gain insights into problem frequency and criticality. Consistent issues include production scheduling, planning and control, efficiency, and lead time reduction. The criticality analysis enables ETO companies to understand problem implications and prioritise decision-making and resource allocation. Addressing the significant criticality problems will minimise negative consequences on ETO project success. The cooperation between industry and academia is crucial for bridging the gap between theoretical knowledge and practical implementation. Continuous monitoring and analysis of criticality indices will aid ETO companies in identifying emerging issues and make informed decisions throughout projects. Balancing high and low criticality problems is crucial to avoid unexpected complications. By understanding ETOs problem criticality and considering Industry 4.0 paradigms, ETO companies can develop targeted strategies for risk reduction and workflow resilience. Effective communication and collaboration within ETO companies and with external stakeholders are vital. Continuous investment in research and development, employee training, and advanced technology adoption ensures long-term success. In addition, fostering solid collaborations between industry and academia addresses critical challenges more effectively. This research finding suggests that ETO companies should improve resource efficiency, adopt innovative management methods, and optimise workflows. Methods such as the implementation of Lean Management principles, which are proven to enhance production performance by improving quality, reducing costs and shortening production times [28], could further augment ETO companies' performance. In addition, continuous monitoring and analysis of criticality indices will aid in identifying emerging issues and making more informed decisions throughout the ETO project lifecycle. As a future proposal, it is recommended to explore the development of a model proposal aimed at improving or optimising the shop floor management process to mitigate the identified problems. This proposal would involve the design, implementation, and validation of a comprehensive framework that integrates resource allocation, workflow optimisation, and technology adoption, specifically tailored for the unique challenges faced by ETO companies. This model could facilitate continuous improvement and drive sustainable growth within the ETO industry by offering practical guidelines and actionable strategies.

References

[1] Jiang, C., Xi, J.T. (2019). Dynamic scheduling in the engineer-to-order (ETO) assembly process by the combined immune algorithm and simulated annealing method, *Advances in Production Engineering & Management,* Vol. 14, No. 3, 271-283, doi: 10.14743/apem2019.3.327.

[2] Brachmann, R., Kolisch, R. (2021). The impact of flexibility on engineer-to-order production planning, *International Journal of Production Economics,* Vol. 239, No. 108183, doi: 10.1016/j.ijpe.2021.108183.

[3] Rother, E.T. (2007). Systematic literature review X narrative review, *Acta Paulista de Enfermagem,* Vol. 20, No. 2, 5-6, doi: 10.1590/S0103-21002007000200001.

[4] Xiao, Y., Watson, M. (2019). Guidance on conducting a systematic literature review, *Journal of Planning Education and Research,* Vol. 39, No. 1, 93-112, doi: 10.1177/0739456X17723971.

[5] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *International Journal of Surgery,* Vol. 88, Article No. 105906, doi: 10.1016/j.ijsu.2021.105906.

[6] McCutcheon, K., Lohan, M., Traynor, M., Martin, D. (2015). A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education, *Journal of Advanced Nursing,* Vol. 71, No. 2, 255-270, doi: 10.1111/jan.12509.

[7] Turnbull, D., Chugh, R. Luck, J. (2021). Learning management systems: A review of the research methodology literature in Australia and China, *International Journal of Research & Method in Education,* Vol. 44, No. 4, 164-178, doi: 10.1080/1743727X.2020.1737002.

[8] Raharjo, T., Purwandari, B. (2020). Agile project management challenges and mapping solutions: A systematic literature review, In: *Proceedings of the 3rd International Conference on Software Engineering and Information Management,* Sydney, Australia, 123-129, doi: 10.1145/3378936.3378949.

[9] Duda, J., Macioł, A., Jędrusik, S., Rębiasz, B., Stawowy, A., Sopińska-Lenart, M. (2021). Quick response manufacturing for high mix, low volume, high complexity manufacturers, *Management and Production Engineering Review,* Vol. 12, No. 1, 72-84, doi: 10.24425/mper.2021.136873.

[10] Papinniemi, J., Fritz, J., Hannola, L., Denger, A., Lampela, H. (2014). Lifecycle-based requirements of product-service system in customer-centric manufacturing. In: Fukuda, S., Bernard, A., Gurumoorthy, B., Bouras, A. (eds.), *Product lifecycle management for a global market, PLM 2014, IFIP Advances in information and communication technology,* Vol. 442. Springer, Berlin, Heidelberg, 435-444, doi: 10.1007/978-3-662-45937-9_43.

[11] Sjøbakk, B., Bakås, O., Bondarenko, O., Kamran, T. (2015). Designing a performance measurement system to support materials management in engineer-to-order: A case study, *Advances in Manufacturing*, Vol. 3, 111-122, doi: 10.1007/s40436-015-0109-2.

[12] Husejnagić, D., Sluga, A. (2015). A conceptual framework for a ubiquitous autonomous work system in the Engineer-To-Order environment, *International Journal of Advanced Manufacturing Technology,* Vol. 78, 1971-1988, doi: 10.1007/s00170-015-6798-7.

[13] Cigolini, R., Gosling, J., Iyer, A., Senicheva, O. (2020). Supply chain management in construction and engineer-to-order industries, *Production Planning & Control,* Vol. 33, No. 9-10, 803-810, doi: 10.1080/09537287.2020.1837981.

[14] Villar-Fidalgo, L., Espinosa Escudero, M.d.M., Dominguez Somonte, M. (2019). Applying kaizen to the schedule in a concurrent environment, *Production Planning & Control,* Vol. 30, No. 8, 624-638, doi: 10.1080/09537287.2019.1566281.

[15] Cannas, V.G., Pero, M., Pozzi, R., Rossi, T. (2018). An empirical application of lean management techniques to support ETO design and production planning, *IFAC-PapersOnLine,* Vol. 51, No. 11, 134-139, doi: 10.1016/j.ifacol.2018.08.247.

[16] Zheng, P., Xu, X., Yu, S., Liu, C. (2017). Personalized product configuration framework in an adaptable open architecture product platform, *Journal of Manufacturing Systems*, Vol. 43, Part 3, 422-435, doi: 10.1016/j.jmsy.2017.03.010.

[17] Vaagen, H., Ballard, G. (2021). Lean and flexible project delivery, *Applied Sciences,* Vol. 11, No. 19, Article No. 9287, doi: 10.3390/app11199287.

[18] Gosling, J., Towill, D.R., Naim, M.M., Dainty, A.R.J. (2015). Principles for the design and operation of engineer-to-order supply chains in the construction sector, *Production Planning & Control,* Vol. 26, No. 3, 203-218, doi: 10.1080/09537287.2014.880816.

[19] Gössinger, R., Plitt, J. (2019). Order release in ETO systems - A basic optimization-based approach, *IFAC-Papers OnLine,* Vol. 52, No. 13, 2207-2212, doi: 10.1016/j.ifacol.2019.11.533.

[20] Xie, W., Nelson, B.L., Barton, R.R. (2014). Statistical uncertainty analysis for stochastic simulation with dependent input models, In: *Proceedings of the Winter Simulation Conference 2014*, Savannah, USA, 674-685, doi:

10.1109/WSC.2014.7019931.

[21] Fortes, C.S., Tenera, A.B., Cunha, P.F. (2023). Engineer-to-order challenges and issues: A systematic literature review of the manufacturing industry, *Procedia Computer Science,* Vol. 219, 1727-1734, doi: 10.1016/j.procs.2023.01.467.

[22] Sylla, A., Vareilles, E., Coudert, T., Aldanondo, M., Geneste, L., Beauregard, Y. (2017). ETO bid solutions definition and selection using configuration models and a multi-criteria approach. In: *Proceedings of 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM),* Singapore, 1833-1837, doi: 10.1109/IEEM.2017.8290208.

[23] Formoso, C., Tommelein, I.D., Saurin, T.A., Koskela, L., Fireman, M., Barth, K., Bataglin, F., Viana, D., Coelho, R., Singh, V., Zani, C., Ransolin, N., Disconzi, C. (2021). Slack in construction - Part 1: Core concepts, In: *Proceedings of 29th Annual Conference of the International Group for Lean Construction (IGLC)*, Lima, Peru, 187-196, doi: 10.24928/2021/0183.

[24] Theoharidou, M., Kotzanikolaou, P., Gritzalis, D. (2009). Risk-based criticality analysis, In: Palmer, C., Shenoi, S. (eds.), *Critical infrastructure protection III. ICCIP 2009, IFIP Advances in information and communication technology,* Vol. 311, Springer, Berlin, Heidelberg, Germany, 3-13, doi: 10.1007/978-3-642-04798-5_3.

[25] Leksic, I., Stefanic, N., Veza, I. (2020). The impact of using different lean manufacturing tools on waste reduction, *Advances in Production Engineering & Management,* Vol. 15, No. 1, 81-92, doi: 10.14743/apem2020.1.351.

[26] Larsson, D., Ratnayake, R.M.C. (2021). Implementation of 5S+S for knowledge work in engineering projects. In: In: Dolgui, A., Bernard, A., Lemoine, D., von Cieminski, G., Romero, D. (eds.*), Advances in production management systems, Artificial intelligence for sustainable and resilient production systems. APMS 2021. IFIP Advances in information and communication technology,* Vol. 630, Springer, Cham, Switzerland, 645-655, doi: 10.1007/978-3-030-85874-2_70.

[27] Bejlegaard, M., Sarivan, I.-M., Waehrens, B.V. (2021). The influence of digital technologies on supply chain coordination strategies. *Journal of Global Operations and Strategic Sourcing,* Vol. 14, No. 4, 636-658, doi: 10.1108/jgoss-11-2019-0063.

[28] Vorkapić, M., Radovanović, F., Ćoćkalo, D., Đorđević, D. (2017). Applicability of the lean concept to the management of small-scale manufacturing enterprises in Serbia, *Tehnički vjesnik – Technical Gazette,* Vol. 24, No. 6, 1929-1934, doi: 10.17559/TV-20150807194942.

APEM journal

Advances in Production Engineering & Management
Volume 18 | Number 2 | June 2023 | pp 199–210
https://doi.org/10.14743/apem2023.2.467

ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Ranking dominant losses in small and medium-sized enterprises (SMEs) in the context of the lean concept application

Kondić, V.[a]**, Maglić, L.**[b]**, Runje, L.**[c]**, Marić, D.**[b,*]

[a]University North, University Center Varaždin, Croatia
[b]Mechanical Engineering Faculty in Slavonski Brod, University of Slavonski Brod, Slavonski Brod, Croatia
[c]Ministarstvo obrane Republike Hrvatske, Croatia

A B S T R A C T

The Lean concept was devised in large business systems and is tailored to this way of conducting business. It is a set of principles, techniques and procedures used to identify and eliminate losses within processes. The results of applying this concept are impressive. Western businesses are delighted with the success of large enterprises that have implemented or have begun to implement the Lean concept. Considering the structures of business systems in transitional and EU countries, a question has arisen as to whether it is possible to apply the Lean concept to small and medium-sized enterprises, as these account for more than 99 % of all business systems. The research which was conducted with the goal of designing a suitable model for the implementation of the Lean concept in small to medium-sized enterprises was based on an analysis of the essential elements of this concept. This article presents part of the conducted research that refers to analysis of losses and identification of the dominant losses according to the opinions of real sector experts and scientists from the academic community. The results of this research were used to define procedures for the elimination of major losses and design a final model for the implementation of the Lean concept in small and medium-sized enterprises.

A R T I C L E I N F O

Keywords:
Manufacturing;
Small and medium-sized enterprises (SMEs);
Lean manufacturing;
Dominant losses;
Ranking;
Analysis of losses;
Elimination of losses

**Corresponding author:*
dmaric@unisb.hr
(Marić, D.)

Article history:
Received 3 June 2022
Revised 6 June 2023
Accepted 19 June 2023

1. Introduction

Worldwide experience has shown that the Lean concept can be successfully applied to improve the processes and overall functioning of large corporate systems, and the results of this application have been impressive. The methodology was first developed in a large Japanese company (TPS-Toyota). It was then disseminated to large companies in the United States and Europe. It was developed and adapted for large companies. As with other successful large system solutions, the Lean concept was gradually adopted by small and medium-sized companies. This has been particularly pronounced more recently, when this methodology has been more intensively researched and applied in Europe and other regions where corporate systems are, on average, much smaller than in the United States [1, 2].

The lean concept envisions products that adapt to the actual needs and expectations of customers, i.e., meet requirements at the required quality and within the agreed delivery times with minimal use of resources [3, 4]. To achieve these goals, it is necessary to eliminate those activities that do not add value or cause waste due to over-production, delays, unnecessary transportation or movement, warehousing, or the like [5, 6]. The authors of [7] state that companies, for example, have successfully reduced their losses and production of waste by applying the lean concept to identify and eliminate or reduce losses. It is important to emphasise that the Lean concept is the industry standard in the automotive industry and its basis for eliminating losses is increasingly being applied in other manufacturing [8] and service sectors. Authors who have studied the lean concept and its application emphasise its importance and role in identifying losses in processes [9, 10]. Errors in the basic approach to the lean concept, e.g. a partial approach, are often cited as reasons for lower success rates in the application of the lean concept [11-13]. Research by authors [14-18] emphasises that the lean concept is expected to improve the capability of companies and increase value for customers through lower prices and improved quality.

The key elements for the success of the lean concept in small and medium enterprises are the efficient identification of processes in relation to their business results, the determination of the top management, the application of the general principles of the lean concept, the reduction of losses, the use of modern lean techniques, continuous improvement, and a good conceptual structure tailored to the particular company [19, 23]. It is important to note that the entire process must begin with questions that identify where resources (time, materials, and the like) are being used and how waste can be reduced or eliminated. This results in the need to identify and rank the dominant losses to finally have an approach for designing an ultimate model for implementing the lean concept in small and medium enterprises [1].

The following chapters briefly describe the most important losses in SMEs, the research and analysis conducted, and the ranking of the losses. Finally, the most important guidelines and actions to eliminate the main losses are recommended to the managers in the production processes. Overview of post-implementation activities.

2. Losses in small and medium-sized enterprises

2.1 Defining losses

Losses or wastes are essentially anything that customers are not willing to pay for. The authors of the article define losses (waste) as all actions in a process that are not necessary for the successful completion of the process. Furthermore, these authors state that after such losses are eliminated, only those operations and activities (called 'value added') remain in the process that are necessary for the successful delivery of such a product or service to customers.

The Japanese word *muda* is translated as 'waste' or 'losses' in most other languages, but the core of the word has a much deeper meaning. The work that takes place in a process is realised through discrete stages or operations. In each of these process stages, input elements are transformed into output elements, either adding or not adding value to the product or service.

Losses are manifested in errors that need to be corrected, in the production of products for which there is no demand, in excessive inventory, and in residual products that are produced unnecessarily. It can also include activities that are not required in the processes, unnecessary movement of employees, unplanned movement of products from one workstation to another, waiting for work, documentation, inspection, handover, or any other form of delay, including complaints from or refunds to customers or product end users.

Taiichi Ohno [24] classified losses in production processes into seven categories: Over-production, unnecessary inventory, defects, unnecessary/superfluous movement, improper processing, waiting, and transportation. Subsequent scientific literature has defined another category of losses, namely losses due to inadequate utilisation of human potential. So, when we analyse losses, we refer to these 7+1 types of losses.

Today, environmental protection as part of the required business efficiency is gaining more importance than in the past. This is because customers and end users are looking for environmentally friendly products and services due to the increasing scarcity of renewable resources and the imposition of (mandatory) regulations. Therefore, this article includes losses attributable to environmental, energy, and safety management systems (green losses) [25].

2.2 The place and role of small and medium-sized enterprises in the economy

For more than two decades, small and medium-sized enterprises have been the engine of economic development. They hire new employees, adapt to changes more easily than large companies, require fewer corporate resources, have shorter and faster communication channels, etc. But their significant role in generating income and investing in fixed assets should not be overlooked either.

In most countries, the size of a company is determined by legislation. Three criteria are considered, total assets, total income, and the number of employees, two of which must reach certain values for a company to be classified in a certain size category. According to the legislation and recommendations of the Commission of the European Union, the main criteria for defining small and medium-sized enterprises are the following [1]:

- Medium-sized enterprises are enterprises with fewer than 250 employees and an annual turnover of less than 50 million euros and/or an annual balance sheet total not exceeding 43 million EUR;
- Small enterprises are enterprises that employ fewer than 50 people and/or whose annual balance sheet total does not exceed 10 million euros; and
- Micro enterprises are enterprises that employ fewer than 10 employees and/or have an annual turnover or annual balance sheet total of less than 2 million euros.

Thus, small and medium enterprises account for 99.7 % of the total number of enterprises in Croatia (92.2 % are micro enterprises, 6.3 % are small and 1.2 % are medium enterprises). They account for 68.3 % of total employment (1.03 million), which is higher than the EU average, and generate added value of 20.5 billion euros (59 %), which is equal to the EU average, and 51 % of GDP [1]. Small and medium-sized enterprises create the most jobs. The fundamental characteristic of small and medium-sized enterprises is their ability to adapt quickly to changing economic conditions.

3. Research implementation

3.1 Research sample selection

Since the research was partially conducted in small and medium enterprises of the real sector, it was necessary to define a meaningful sample eligible for this research. This was ensured by an appropriate selection of companies, taking into account their activities and their international classification (European Accreditation Classification – EAC). It is important to point out that the studied sample of small and medium enterprises is from the industrial sector, i.e. manufacturing.

The researcher's assessment and the opinion of experts on the lean concept led to an approximate number of SMEs in the Republic of Croatia that apply the lean concept. The sample size was 36 SMEs, which corresponds to 34 % of the studied companies. According to the form and characteristics of the sample, it was a quota sample, i.e. an intentional sample, which means that the researchers of this article selected the companies to be studied. One to two respondents from each of the 26 selected small and medium enterprises participated in the survey, making the total number of respondents 30. The sample included companies that were 'reachable' at the time this research was conducted, were willing to provide the required responses, and were at various stages of implementing the lean concept. The criteria used to select the study participants from the selected small and medium enterprises were their knowledge of the Lean concept and their position in the company [1].

In this research, an appropriate opinion of scientists from the academic community was obtained. Care was taken to select scientists from the Republic of Croatia and from other countries.

The aim was to compare the opinions of experts from practice and academia on the problem of losses in SMEs. A total of 85 surveys were sent out, 30 of which were answered within a reasonable period of time. The opinions of scientists from academic communities outside the Republic of Croatia were also included in this research. Several times, about 220 surveys were sent out, of which 30 were returned within the time limit. The criterion of knowledge of the Lean concept and experience in theoretical and practical work with its implementation was used in the selection of scientists [1].

It is assumed that the sample size is significant for the field in which the research was conducted. Since the opinion of experts from all over the world was also obtained, it is assumed that the results can be generalized to countries that have a similar economic structure as the Republic of Croatia, but taking into account their specifics.

3.2 Research methodology and ranging losses

Ranging losses in the real sector

Research has been conducted in both the real sector and the academic community to classify the prevailing losses. A method based on Spearman's Rank correlation coefficient was developed to rank the losses in the private sector [26]. The real sector ranking was done using a semi-structured interview. Participants assigned an appropriate rank to each factor based on their opinion: rank 1 (the most important) to rank 9 (the least important). Factors with equal influence were also ranked in the same way. Respondents could assign the same ranks to different factors.

Respondents could assign the same ranks to different factors. For these reasons, the ranks had to be redesigned. Therefore, factors with the same rank received a new rank that corresponded to the mean value of the rank that the factors shared among themselves. Multiple factors with the same rank value were assigned a related rank. The average rank was determined as the arithmetic mean of the ranks that the factors would have received if they had not been related ranks. With the rank table redesigned, the null and alternative hypotheses for the Spearman's Rank correlation coefficient were as follows:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

The probability of a type I error was set an extremely high significance limit, i.e. α = 0.01. The significance limit α denotes an area of rejection of the $H0$ hypothesis. Calculation of Spearman's Rank correlation coefficient ρ is suitable if the number of pairs in the sequence is less than or equal to 30, and is calculated using Eq. 1.

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{9} (R^{(1)} - R^{(2)})^2}{N \cdot (N^2 - 1)} \tag{1}$$

where R is the squared ranks' difference of the corresponding variable value pair, and N is the number of influencing factors (N = 9).

The verification of the statistical value of ρ was performed using the hypothesis test for the significance of the correlation coefficient tr. The statistical value for ρ was verified using the t-distribution, whose value was calculated with Eq. 2.

$$t_r = \frac{\rho}{\sqrt{\frac{1 - \rho^2}{N - 2}}} \tag{2}$$

where ρ is Spearman's Rank correlation coefficient (ρ = 0.984).

Since the calculated value t_r (15.04) was higher than the tabular value t_t (2.99), it was assumed that the correlation coefficient was significant. With this statement, a one-sided and positive correlation was confirmed. Since the value of Spearman's Rank correlation coefficient reached 1, it could be confirmed with high reliability that the newly designed ranking tables were suitable for initial ranking and could be used for further analysis.

Since there were several equal ranks in the range of losses defined by the experts, the expression for the Spearman's Rank correlation coefficient had to be corrected. The corrected rank correlation coefficient is called Kendall's concordance coefficient and is usually referred to as Kendall's W in the literature. By its logic, the concordance coefficient W tests the relationship between the reviewer's actual concordance and the maximum possible concordance. Eq. 3 was used for the calculation

$$W = \frac{12 \sum S^2}{m^2(N^3 - N) - m \sum_{i=1}^{m} T_i} \tag{3}$$

where m is the number of interlocutors conducting significance ranking factors (m = 30), T_i is the sum of correction factors for related ranks, and S^2 is the sum of squared sum of the rank deviations of all respondents' and the sum of the averages.

The significance of Kendall's coefficient of concordance (W) was performed by applying the χ^2-test, as defined by Eq. 4. This required that the calculated value of this value is greater than that in the table.

$$\chi_r^2 = m \cdot (N - 1) \cdot \mathrm{W} \tag{4}$$

Since the ratio between the calculated and tabulated values was 46.56 > 20.1, the hypothesis test regarding the agreement of the opinions of the respondents showed that the ranks were interdependent, which means that the hypothesis of the agreement of the ranks of the respondents could be accepted. The calculation, i.e. the determination of the strength of the influence of all factors, was performed with the help of the dominance coefficient ϕ, which was calculated with the Eq. 5.

$$\phi = \frac{\sum_{i=1}^{m} s_{ij} \cdot \tau_i}{\sum_{i=1}^{m} \sum_{j=1}^{k} \tau_i \, s_{ij}} \tag{5}$$

where τ represents the coefficient of participating experts' competences. The values of the dominance coefficient are shown in Fig. 1.

Ranging losses by scientists

A suitable questionnaire was developed for the purpose of this research. It was designed according to the previous selection of variables and allowed the assessment of the respondents involved. The form contained 9 statements with 5 response options that corresponded to the Likert scale, allowing the expression of positive or negative attitudes towards each statement. In their responses, respondents expressed the degree of their agreement or disagreement with the position expressed in the statement (disagree at all, agree, do not know/neutral, agree, and strongly agree). After the initial evaluation, a weighted average calculation was performed for each participant, from which an individual ranking was determined for all 9 variables. These ranks were then used in the subsequent analysis. The weighted average, also referred to as the weighted arithmetic mean, is the average in which a weight (the weighting factor) is assigned to each quality item. The difference with the arithmetic mean is that the data that are averaged are not considered important. In this way, these calculated weights determine the relative average importance of each quantity [1].

Table 1 and Fig. 2 show the ranking of losses of scientists within the Republic of Croatia according to the weighted average. An identical procedure was performed with scientists outside the Republic of Croatia. Table 2 and Fig. 3 show their ranking of losses.

The statements offered to the participants were related to the impact of losses on business results, and for each of the following 9 statements, the participants' opinion was sought on which aspect is the main cause of losses in the production process:

Statement 1 – 'Over-production' is the dominant cause of losses.
Statement 2 – 'Inventory' is the dominant cause of losses.
Statement 3 – 'Transportation' is the dominant cause of losses.
Statement 4 – 'Waiting' is the dominant cause of losses.
Statement 5 – 'Excess motion' is the dominant cause of losses.

Statement 6 – 'Defects' are the dominant cause of losses.
Statement 7 – 'Over-processing' is the dominant cause of losses.
Statement 8 – 'Utilisation of human potential' is the dominant cause of losses.
Statement 9 – 'Green losses' are the dominant cause of losses.

3.3 Brief description of losses

Losses related to over-production

Over-production means producing more, earlier, or faster than is required for the next stage of the production process. It is neglected by companies that view over-production and larger inventories as a safety rather than a loss. Over-production is often the result of production planning based on sales forecasts rather than specific orders. In addition, over-production can be the result of poor communication and business relationships with suppliers, the use of high performance hardware, extensive labour, long lead times, poor business decisions, and a host of other reasons. In both the technical literature and academia, this is considered a major cause of losses. The cost of initialising machinery and production processes beyond what is required is often a hidden loss and must be accounted for. Combined with the additional cost of inventory or the inability to sell additional inventory after a certain time, this can lead to further losses [27].

The lean concept is based on the pull principle of production, which involves manufacturing products according to customer orders. This means that production must be adjusted to demand, i.e., the goal should be to produce only what is needed when a customer demands it, according to the 'just in time' principle, while many companies use the 'just in case' principle. The statement for the survey participants was: 'Over-production' as a loss represents a dominant impact on the efficiency of the production process.

Losses related to inventory

Accumulating unnecessary inventories of materials, work in process, tooling, or finished goods creates unnecessary costs. It is important to note that unnecessary inventory adds no value to the final product and represents a net cost to the manufacturer until it is sold to a customer. This cost is a significant problem for SMEs because unnecessary inventory creates additional space problems, additional moves and transportation, and the possibility of damage or reduction in the quality or value of the products. The statement for survey respondents was that 'inventories' as a loss represent a dominant impact on the efficiency of the production process.

Losses related to transportation

Transportation is the movement of materials and products from one place to another. This is a loss because it does not add value to the product and does not impose a cost that customers are willing to pay for. For manufacturers, it is an unnecessary cost because they must pay employees who are directly or indirectly involved in internal or external transportation, and they must also provide adequate transportation and appropriate equipment and supplies. Regardless of whether transportation is optimally set up or integrated into the technological process, it always represents a loss [28]. In this sense, the lean concept requires an analysis of all technological processes and the elimination or shortening of unnecessary transportation. The statement for the survey participants was: 'Transport' as a loss has a dominant influence on the efficiency of the production process.

Losses related to waiting

Waiting times are very often the result of machine breakdowns, supply problems, poor material and capacity planning, poor design of technological procedures used in the production process, poor structuring of documented information and poor control structures, and so on. Waiting time is the time lost due to slow or stopped production in a single part of the production process while the previous step is being completed. A classic example of waiting time is any production line with multiple orders or operations. If there is a different technological time between two operations, i.e. if the processing time of the second operation is longer than that of the first or vice versa, waiting times or losses occur. Operations that require more time must be made more efficient or production must be balanced by measuring and calculating with production rates. Losses due to waiting lead to a disruption in the flow of work, which is one of the main principles of lean production [29]. The statement for the survey participants was: 'Waiting' as a loss has a dominant influence on the efficiency of the production process.

Losses related to excess motion at workplaces

Any movement of workers in the workplace that is not directly related to value creation is unproductive and represents a loss. For example, by observing a worker in a processing system, it is possible to determine when value is actually being added. In theory, this is a very short period of time, sometimes only a few seconds. The other movements are actions that do not add value, such as lifting or lowering in the position where processing is taking place. The few seconds lost in such an operation due to unnecessary employee movements become minutes and eventually hours due to repetitive activities. Excessive movements in the workplace are often the result of poor ergonomic solutions in the machines, tools or equipment themselves. The statement for the survey participants was: 'Excessive movement' has a dominant influence on the efficiency of the production process as a loss.

Losses related to defects

Defects (scrap) in production can occur in items, assemblies, or finished products. It results from quality control measures that identify a specific nonconformance of a product to requirements, specifications, or contracts. For non-conforming products, repairs (rework) of deterioration must be made. If for any reason it is not possible to make such repairs, or if a customer is unwilling to purchase the product at a reduced price, the manufacturer must scrap the product as defective. All actions taken to fix a nonconforming product represent additional costs because time and materials are lost in the manufacturing process. However, the greatest cost to manufacturers comes from nonconforming products reaching the customer. Sometimes these costs can be much higher than expected [30]. This can lead to high reclamation or modification costs or simply the loss of customers. Non-conforming products result from frequent design changes, machine setup errors, decreasing operator concentration, errors in production documentation, and the like. The statement for survey participants was that 'defects' represent a dominant impact on the efficiency of the production process as a loss [31, 32].

Losses related to over-processing

Over-processing adds value to a product that customers have neither asked for nor want to pay for. Such activities represent an unnecessary cost to manufacturers. This loss comes in the form of lost time for employers and use of resources. Over time, these are costs that can significantly reduce the efficiency and effectiveness of a given process. Over-production can result from unclear standards and specifications or poor technical documentation. Many manufacturers strive to produce the highest quality product possible, but do not know what truly adds value to a product or what is essential to the end user. Examples of over-processing include painting surfaces that will never be seen or exposed to corrosive processes, or polishing and nickel plating surfaces that are not required by the customer. The statement for survey respondents was, 'Over-machining' as a loss represents a dominant impact on the efficiency of the production process.

Losses related to the utilisation of human potential

When existing knowledge and entrepreneurial know-how are not fully utilised, this harms companies in various ways: Loss of efficiency, unnecessary new hires, demotivated employees who lack recognition, etc. Employees remain underutilised, and companies fail to realise the full potential of their workforce. Employees are a company's most valuable resource, and without their commitment and loyalty, companies cannot be as competitive as they can be in the marketplace. It is very important that the contribution of all employees is recognised [1]. Most companies do not allow their employees to participate in the production process beyond what is required because they fear that they will become overqualified, demand higher wages, or leave for another company that can offer them better conditions as compensation for their newly acquired knowledge and experience. The statement for the survey participants was that 'the non-utilisation of human potential' represents as a loss a dominant impact on the efficiency of the production process.

It must be emphasised that this loss is the most difficult to identify and quantify and, in this context, the most difficult to eliminate.

Losses related to environmental protection

As stated in the introductory part of this article, losses in the production process and in other processes of any system are also associated with losses related to the energy management system, environmental protection, and workplace safety. Seven types of losses have been identified, grouped under the title 'Green Losses'. They are related to green production, minimising process waste and pollution from processes, and from the design and execution of products and services. The statement for survey participants was that 'Green losses' represent a dominant impact on the efficiency of the production process as a loss.

3.4 Results of loss ranging

The results of loss ranging by competent experts from the field are shown in Fig. 1, and the results of loss ranging by scientists within the Republic of Croatia (ZRH) are shown in Table 1 and Fig. 2. The results of loss ranging by scientists outside the Republic of Croatia (ZIRH) are presented in Table 2 and Fig. 3.

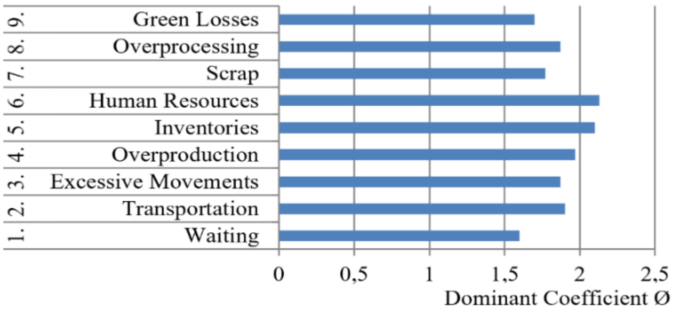


Fig. 1 A presentation of loss ranging as defined by real sector professionals

Table 1 Loss ranging on the basis of weighted sverages (Croatia)

Ordinal Number	Loss Type -Statement-	Degree of Agreement* SD	D	N	A	SA	Weighted Average	Range
1.	Over-production	1	12	6	10	1	1.97	7
2.	Inventories	0	8	8	11	3	2.10	8
3.	Transportation	0	9	6	13	2	1.90	6
4.	Waiting	0	1	4	18	7	1.60	1
5.	Excessive Movements	1	4	11	13	1	1.87	4
6.	Scrap	0	4	2	15	9	1.77	3
7.	Over-processing	0	5	7	15	3	1.87	5
8.	Utilisation of Human Resources	1	4	6	13	6	2.13	9
9.	Green Losses	0	5	18	5	2	1.70	2

* SD-Strongly Disagree; D-Disagree; UN-Unknown/Neutral; A-Agree; SA-Strongly Agree

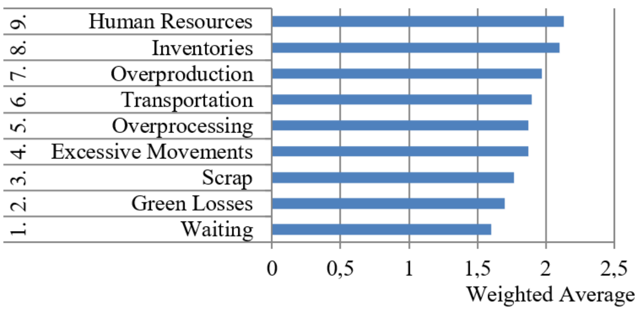


Fig. 2 A presentation of loss ranging as defined by experts from within Croatia

Table 2 Loss ranging on the basis of weighted averages (outside Croatia)

Ordinal Number	Loss Type -Statement-	Degree of Agreement*					Weighted Average	Range
		SD	D	UN	A	SA		
1.	Over-production	0	4	8	14	4	1.93	6
2.	Inventory	2	6	6	10	6	2.47	9
3.	Transportation	0	8	4	17	1	1.63	3
4.	Waiting	0	0	3	16	11	1.56	2
5.	Excessive Movements	2	4	5	15	4	1.97	5
6.	Scrap	1	5	6	12	6	2.23	7
7.	Over-processing	1	6	0	18	5	1.63	4
8.	Utilisation of Human Resources	3	6	5	11	5	2.43	8
9.	Green Losses	6	2	21	1	0	1.43	1

* SD-Strongly Disagree; D-Disagree; Unknown/N-Neutral; A-Agree; SA-Strongly Agree

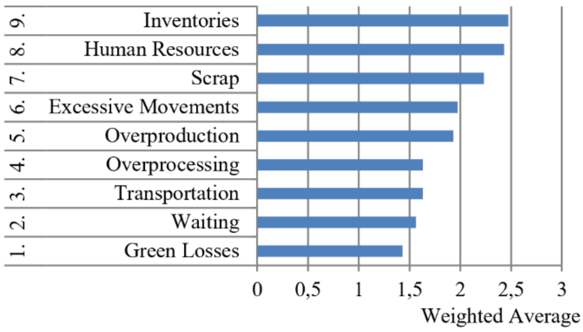


Fig. 3 A Presentation of loss ranging as defined by experts from outside Croatia

Statistical data verification

Because the data describing the losses are ordinal in nature, the analysis involves ranking the data rather than the data themselves. In nonparametric statistics, the Kruskal-Wallis test is most commonly used to compare the ranks of losses from three or more samples. In this study, three samples were used. The first sample refers to the private sector (RS), the second to academic scientists from the Republic of Croatia, and the third to academic experts from outside the Republic of Croatia. The Kruskal-Wallis test essentially tests the hypotheses that have been established. It is a test of an analysis of variance, except that instead of numerical measurement data (continuous variables), rankings (discrete variables) are used [1]. A test of statistic H is created by the Eq. 6.

$$H = \frac{12}{N(N+1)} \sum \frac{T_i}{N_i} - 3(\mathrm{N}+1) \tag{6}$$

where N is the total number of observations, and T_i is the sum of the ranks in a single sample. If the samples are large enough (in this case the samples are considered large enough because each sample contains 5 results), H has the same distribution as the *HI*-squared, so its significance can be read in the χ^2 Table with $k-1$ denoting the degree of freedom and the level of significance α. The results of the Kruskal-Wallis test for the individual samples are given in Tables 3, 4 and 5 for a significance of 5 % (α = 0.05).

Table 3 Analysis of losses within the RS sample

RS			
Test Statistic H	Value of p	Test Statistic H Custom Value	Value of p
43.40	0.000	43.43	0.000

Table 4 Analysis of losses within the 'Croatia' sample

Croatia			
Test Statistic H	Value of p	Test Statistic H Custom Value	Value of p
32.08	0.000	35.63	0.000

Table 5 Analysis of losses within the ZIRH sample

Outside Croatia			
Test Statistic H	Value of p	Test Statistic H Custom Value	Value of p
39.38	0.000	43.40	0.000

Before determining the final ranking of losses, it is necessary to determine whether there is a statistically significant difference between the three samples in terms of perceptions of the intensity of the impact of these losses on the overall efficiency of small and medium-sized enterprises. To test this, the hypotheses $H0$ and $H1$ were established.

$H0$ – there is no significant statistical difference between RS, 'Croatia' and 'Outside Croatia' in terms of perception of the impact of losses on the overall efficiency of small and medium enterprises.

$H1$ – there is a significant statistical difference between RS, 'Croatia' and 'Outside Croatia' in terms of perception of the impact of losses on the overall efficiency of small and medium enterprises.

The results of the Kruskal-Wallis test in relation to the perception of the impact of losses on the overall efficiency of small and medium-sized enterprises on the part of RS, 'Croatia' and 'Outside Croatia' are presented in Table 6. These results indicate whether the null hypothesis can be rejected, i.e. an alternative is accepted at the selected significance level. Specifically, these test results show that the ranks defined by the real sector, scientists from the Republic of Croatia and scientists from outside the Republic of Croatia are not statistically significantly different in terms of the perception of the impact of losses on the overall efficiency of SMEs (p-value is 0.084, i.e. $p > 0.05$). From this it could be concluded that it is statistically correct to accept the $H0$ hypothesis.

Table 6 Results of the Kruskal-Wallis test for the Three Samples (RS, Croatia and Outside Croatia)

RS, Croatia and Outside Croatia			
Test Statistic H	Value of p	Test Statistic H Custom Value	Value of p
13.90	0.084	14.06	0.080

From the analysis and comparison of the individual ranks (see Tables 1 and 2), the following can be deduced. When analysing the ranking determined by scientists from the Republic of Croatia (Croatia), a high ranking (ranking number 2) for 'Green Losses' can be observed. In this particular case, 18 participants or 60 % declared themselves neutral/undecided about this type of loss. This means that they did not have a specific opinion about this type of loss. Similarly, 5 or 16.7 % stated that they disagreed, while 16.7 % expressed agreement.

The analysis of the ranking determined by researchers outside the Republic of Croatia (Outside Croatia) showed, as mentioned above, a high ranking (rank 1) for 'Green losses'. In this particular case, 21 participants, or 70 %, stated that they were neutral/undecided about this type of loss, while 5, or 20 %, disagreed with the statement that this is the predominant loss that significantly affects the business efficiency of SMEs. For these reasons, the responses of the participants from 'Croatia' and 'Outside Croatia' regarding 'Green Losses' are excluded from the analysis.

Taking into account the results of the Kruskal-Wallis test and the above statement, the rankings were combined and the final results of the loss ranking (Table 7) were presented. There are three 'significant losses' for SMEs, namely: waiting time, transportation, and excessive movement. There are other losses, but these three are prevalent and require more attention in improving the processes.

Table 7 An analysis of individual ranks and a presentation of the final loss rankings

Code	LOSSES	RANGE 1 'RS'	RANGE 2 'ZRH'	RANGE 3 'ZIRH'	SUM OF RANGES	FINAL RANGE
G_1	Over-production	4	7	6	17	4
G_2	Inventory	5	8	9	22	5
G_3	Transportation	2	6	3	11	2
G_4	Waiting	1	1	2	4	1
G_5	Excessive Movement	3	4	5	12	3
G_6	Scrap	7	3	7	17	4
G_7	Over-processing	8	5	4	17	4
G_8	Utilisation of Human Resources	6	9	8	23	6
G_9	Green Losses	9	2	1	12	-

4. Conclusion

Eliminating and reducing losses is one of the main goals of any improvement system, and this includes the Lean concept. This paper is mainly about identifying the most important loss types from the defined 9 loss types that occur in the production environment. The calculation of the ranks is based on the opinions of the surveyed participants using Spearman's rank correlation coefficient and weighted average method. The results indicate that the top three loss types were identified from the total nine losses. Most respondents indicated that losses related to waiting time, transportation, and excessive movements were the main causes of losses for small and medium enterprises. Statistical analysis of the results confirmed this conclusion. Based on the results of this research, it can be concluded that production managers and professionals responsible for the production process must systematically monitor these three main types of losses so that they can be eliminated or continuously reduced. Elimination of the prevailing losses significantly increases productivity and other indicators of production processes [33, 34].

The activities of those responsible for loss elimination should focus on the following:

- Identification of losses;
- Implementation of the proposed loss elimination systems;
- Training employees in the application of the lean concept to eliminate or reduce losses;
- Measuring losses;
- Establishing liability for losses,
- Analysing risks and opportunities for loss management.

Specific procedures for eliminating losses are based on the nature of the losses, their magnitude, and their priorities. Procedures include continuous training of staff and planning of activities of responsible staff with continuous measurement, monitoring and improvement.

Various lean techniques can be applied to reduce losses in production systems, such as: 5S, Kanban, Andon, production bottleneck analysis, Kaizen, Jidoka, JIT, TPM, VSM, Gemba, SMED, OEE, PDCA, TQM, SMART goals, Continuous production flow, Tact time, Visual Factory, Poka-Yoke, etc. [35, 36]. The application of various lean techniques is essential in all occasions for the identification, analysis and elimination of losses, with emphasis on numerical indicators (Six Sigma).

Well-organised loss management and the lean concept in general help SMEs to withstand unforeseen situations such as a pandemic, political and economic unrest, etc. unforeseen events.

References

[1] Kondić, V. (2010). *Poboljšavanje efikasnosti procesa na načelima Lean koncepcije u malim i srednjim poduzećima,* PhD thesis, University of Slavonski Brod, Slovonski Brod, Croatia.

[2] Stoijljković, V. (2017). ISO 9001:2015 i Lean, *Integracija dva koncepta menadžmenta – ISO 9001:2015 sistem menadžmenta kvalitetom i Lean koncept,* Talija Izdavaštvo, Niš, Serbia.

[3] Piškor, M., Kondić, V. (2010). Lean production kao jedan od načina povećavanja konkurentnosti hrvatskih poduzeća na globalnom tržištu, *Tehnički glasnik,* Vol. 4, No. 1-2, 37-41.

[4] Saad, D.A., Masoud, M., Osman, H. (2021). Multi-objective optimization of lean-based repetitive scheduling using batch and pull production, *Automation in Construction,* Vol. 127, Article No. 103696, doi: 10.1016/j.autcon.2021.103696.

[5] Arunagiri, P., Gnanavelbabu, A. (2014). Identification of major lean production waste in automobile industries using weighted average method, *Procedia Engineering,* Vol. 97, 2167-2175, doi: 10.1016/j.proeng.2014.12.460.

[6] Arunagiri, P., Gnanavelbabu, A. (2014). Identification of high impact lean production tools in automobile industries using weighted average method, *Procedia Engineering,* Vol. 97, 2072-2080, doi: 10.1016/j.proeng.2014.12.450.

[7] Vorkapić, M., Radovanović, F., Ćoćkalo, D., Đorđević, D. (2017). Applicability of the lean concept to the management of small-scale manufacturing enterprises in Serbia, *Tehnički Vjesnik – Technical Gazette*, Vol. 24, No. 6, 1929-1934, doi: 10.17559/TV-20150807194942.

[8] Gošnik, D., Beker, I., Kavčič, K. (2014). Lean six sigma in Slovenian and Serbian manufacturing companies, *International Journal of Industrial Engineering and Management,* Vol. 5, No. 3, 123-130.

[9] Berah, I. (2016). *Startegija unapređenja poslovnih performansi malih i srednjih poduzeća u lancu opskrbe automobilske industrije,* PhD thesis, Univerzitet Union-Nikola Tesla, Beograd, Serbia.

[10] Antosz, K., Stadnicka, D. (2017). Lean philosophy implementation in SMEs – Study results, *Procedia Engineering,* Vol 182, 25-32, doi: 10.1016/j.proeng.2017.03.107.

[11] Shrimali, A.K., Soni, V.K. (2018). A study on the utilization of lean techniques/tools in Indian SMEs, *Production Engineering Archives,* Vol. 20, No. 20, 32-37, doi: 10.30657/pea.2018.20.07.
[12] Ulewicz, R., Kucęba R. (2016). Identification of problems of implementation of Lean concept in the SME sector, *Engineering Management in Production and Services,* Vol. 8, No. 1, 19-25, doi: 10.1515/emj-2016-0002.
[13] Prasanna, M., Vinodh, S. (2013). Lean Six Sigma in SMEs - An exploration through literature review, *Journal of Engineering, Design and Technology,* Vol. 11, No. 3, 224-250, doi: 10.1108/JEDT-01-2011-0001.
[14] Majava, J., Ojanperä, T. (2017). Lean production development in SMEs: A case study, *Management and Production Engineering Review,* Vol. 8, No. 2, 41-48, doi: 10.1515/mper-2017-0016.
[15] Rymaszewska, A.D. (2016). *Rethinking the applicability of lean philosophy- A conceptual and empirical analysis,* PhD thesis, University of Vaasa, Finland.
[16] Rose, A.M.N., Deros, B.M., Rahman, M.N.A., Nordin, N. (2011). Lean manufacturing best practices in SMEs, In: *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management*, Kuala Lumpur, Malaysia, 872-877.
[17] Xu, L.X.X., Wang, F.Y., Lim, R., Toh, M.H., Valliappan, R. (2013). Lean implementation in small and medium enterprises – A Singapore context, In: *Proceedings of the 2013 IEEE International Conference on Industrial Engineering and Engineering Management,* Bangkok, Thailand, 1592-1596.
[18] Kurdve, M., Bellgran, M. (2021). Green lean operationalisation of the circular economy concept on production shop floor level, *Journal of Cleaner Production,* Vol 278, Article No. 123223, doi: 10.1016/j.jclepro.2020.123223.
[19] Shaqour, E.N. (2022). The impact of adopting lean construction in Egypt: Level of knowledge, application, and benefits, *Ain Shams Engineering Journal,* Vol 13, No. 2, Article No. 101551, doi: 10.1016/j.asej.2021.07.005.
[20] April, J., Powell, D., Bart, S. (2010). A new lean change methodology for small & medium sized enterprises, In*: Proceedings of 12th International MITIP Conference on Information Technology & Innovation processes of the enterprises,* Alborg, Denmark, 27-35.
[21] Womack, J.P., Jones, D.T. (2003). *Lean thinking: Banish waste and create wealth in your corporation,* Simon & Schuster, London, United Kingdom.
[22] Kizielewicz, B., Więckowski, J., Shekhovtsov, A., Wątróbski, J., Depczyński, R., Sałabun, W. (2021). Study towards the time-based MCDA ranking analysis – A suppliner selection case study, *Facta Universitatis Series: Mechanical Engineering,* Vol. 19, No. 3, 381-399, doi: 10.22190/FUME210130048K.
[23] Edwin Joseph, R., Kanya, N., Bhaskar, K., Francis Xavier, J., Sendilvelan, S., Prabhahar, M., Kanimozhi, N., Geetha, S. (2021). Analysis on productivity improvement, using lean manufacturing concept, *Materials Today: Proceedings,* Vol. 45, Part 7, 7176-7182, doi: 10.1016/j.matpr.2021.02.412.
[24] Gopal, N., Panchal, D. (2021). A structured framework for reliability and risk evaluation in the mild process industry under fuzzy environment, *Facta Universitatis Series: Mechanical Engineering*, Vol. 19, No. 2, 307-333, doi: 10.22190/FUME201123004G.
[25] Ohno, T. (1988). *Toyota production system, Beyond large-scale production,* First edition, Productivity Press, New York. USA.
[26] Veža, I., Gjeldum, N., Mladineo, M., Babić, B., Čelar, S. (2015). *Analiza postojećeg stanja hrvatskih industrijskih poduzeća – Projektni izveštaj RP 1,* Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia.
[27] Kondić, Ž., Maglić, L., Samardžić, I. (2009). Analysis and ranking of factors impacting application of the 6σ-methodology in small production organizations using the prior factor ranking method, *Tehnički Vjesnik – Technical Gazette,* Vol. 16, No. 2, 17-25.
[28] Conti, R., Angelis, J., Cooper, C., Faragher, B., Gill, C. (2006). The effects of lean production on worker job stress, *International Journal of Operations & Production Management,* Vol. 26, No. 9, 1013-1038, doi: 10.1108/01443570610682616.
[29] Shirouzu, N., Moffett, S. (2004). Toyota closes in on GM, quality concerns also grow, *The Wall Street Journal,* from *https://www.wsj.com/articles/SB109151488383981450*, accessed November 30, 2021.
[30] Kondić, Ž., Maglić, L., Pavletić, D., Samardžić, I. (2018). *Kvaliteta*, University North Varaždin, Mechanical Engineering Faculty in Slavonski Brod – University of Slavonski Brod, Faculty of Engineering – University of Rijeka, Croatia.
[31] Jurczyk-Bunkowska, M. (2021). Tactical manufacturing capacity planning based on discrete event simulation and throughput accounting: A case study of medium sized production enterprise, *Advances in Production Engineering & Management*, Vol. 16, No. 3, 335-347, doi: 10.14743/apem2021.3.404.
[32] Rosin, F., Forget, P., Lamouri, S., Pellerin, R. (2021). Impact of Industry 4.0 on decision-making in an operational context, *Advances in Production Engineering & Management*, Vol. 16, No. 4, 500-514, doi: 10.14743/apem2021.4.416.
[33] Bucko, M., Schindlerova, V., Krupova, H. (2022). Application of lean manufacturing methods in the production of ultrasonic sensor, *Tehnički Vjesnik – Technical Gazette,* Vol. 29, No. 5, 1671-1677, doi: 10.17559/TV-20220421141917.
[34] Gajšek, B., Cajner, H., Butlewski, M., Opetuk, T., Đukić, G. (2022). Bi-objective assignment model for lean order picking in a warehouse, *Tehnički Vjesnik – Technical Gazette,* Vol. 29, No. 1, 293-300, doi: 10.17559/TV-20201002103433.
[35] Leksic, I., Stefanic, N., Veza, I. (2020). The impact of using different lean manufacturing tools on waste reduction, *Advances in Production Engineering & Management*, Vol. 15, No. 1, 81-92, doi: 10.14743/apem2020.1.351.
[36] Sá, S., Ferreira, L.P., Silva, F.J.G., Sá, J.C., Pereira, M.T., Santos, G. (2022). The importance of subcontracting and its relationship with lean philosophy in automotive industry, *International Journal of Industrial Engineering and Management*, Vol. 13, No. 3, 186-193, doi: 10.24867/IJIEM-2022-3-311.

APEM journal
Advances in Production Engineering & Management
Volume 18 | Number 2 | June 2023 | pp 211–224
https://doi.org/10.14743/apem2023.2.468
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

An improved discrete particle swarm optimization approach for a multi-objective optimization model of an urban logistics distribution network considering traffic congestion

Li, K.[a]**, Li, D.**[a,*]**, Ma, H.Q.**[b]

[a]School of Business Administration, Shanghai Lixin University of Accounting and Finance, Shanghai, P.R. China
[b]Human Resources Department, Shanghai Institute of Visual Arts, Shanghai, P.R. China

ABSTRACT

To optimize urban logistics networks, this paper proposes a multi-objective optimization model for urban logistics distribution networks (ULDN). The model optimizes vehicle usage costs, transportation costs, penalty costs for failing to meet time windows, and carbon emission costs, while also considering the impact of urban road traffic congestion on total costs. To solve the model, a DPSO (Discrete Particle Swarm Optimization) algorithm based on the basic principle of PSO (Particle Swarm Optimization) is proposed. The DPSO introduces multiple populations to handle multiple targets and uses a variable neighbourhood search strategy to improve the search ability of particles, which helps to improve the local search ability of the algorithm. Simulation results demonstrate the effectiveness of the proposed model in avoiding traffic congestion, reducing carbon emissions costs, and time penalty costs. The optimization comparison results between DPSO and PSO also verify the superiority of the DPSO algorithm. The proposed model can be applied to real-world urban logistics networks to improve their efficiency, reduce costs, and minimize environmental impact.

ARTICLE INFO

Keywords:
Urban logistics distribution network;
Traffic congestion;
Optimization;
Modelling;
Multi-objective optimization;
Vehicle routing problem (VRP);
Swarm intelligence;
Discrete particle swarm optimization algorithm (DPSO)

**Corresponding author:*
gslidan@lixin.edu.cn
(Li, D.)

Article history:
Received 23 April 2023
Revised 7 June 2023
Accepted 13 June 2023

1. Introduction

As urban populations continue to grow, the demand for urban logistics services has increased significantly. The logistics industry plays a crucial role in the economic development of cities [1-6]. However, the development of urban economies and the increase in private cars have led to increased traffic congestion, which has added pressure to the urban transportation network. Logistics companies must not only deal with traffic congestion during distribution, but also meet increasingly strict time requirements from customers. To meet the growing demand for timely delivery services under complex and unstable urban traffic conditions, logistics companies must find ways to provide efficient and effective distribution services that meet customer needs.

In the process of urban logistics distribution, the shortest distance between two customer nodes is often not the most efficient route, especially under urban traffic conditions where the shortest distance route is often the most congested. To avoid congested sections, shorten delivery times, and respond to customer time needs, logistics companies require an optimized distribution network. Improving the efficiency of logistics service delivery while meeting customer needs and ensuring customer loyalty can not only improve the core competitiveness of enterprises, but also alleviate urban traffic congestion and improve the urban environment.

In addition, environmental pollution caused by urban logistics activities has become a critical issue. Low-carbon logistics plays an important role in the low-carbon development of the environment and provides an essential guarantee for economic and social low-carbon development. Therefore, logistics companies must address the environmental problems caused by urban logistics activities. Green distribution is a necessary consideration for cities to optimize the overall logistics facility layout and distribution network. To optimize urban logistics networks, it is essential to consider both traffic congestion and green distribution, as it can improve the economic benefits of logistics enterprises while reducing the carbon emissions generated in the process of logistics distribution.

This paper proposes an urban logistics network optimization model that considers the impact of traffic congestion on urban logistics distribution. The model aims to minimize vehicle usage costs, transportation costs, time penalty costs, and carbon emission costs. To solve the model, we introduce an improved discrete particle swarm optimization (DPSO) algorithm that introduces multiple populations to handle multiple targets and uses a variable neighbourhood search strategy to enhance the search ability of particles and improve the local search ability of the algorithm. Simulation experiments verify the effectiveness of the proposed model and the superiority of the DPSO algorithm.

2. Literature review

The vehicle routing problem (VRP) is a fundamental problem in the optimization of urban logistics networks. Since its inception in 1945, numerous research studies have been published [7-15]. Gulczynski *et al.* [16] proposed a vehicle routing problem based on batch delivery and developed a heuristic method to solve the problem. Contardo and Martinelli [17] studied the multi-site vehicle routing problem with capacity and routing length constraints and designed a new exact algorithm to solve the problem. Bertazzi and Secomandi [18] focused on vehicle routing problems with random demand and replenishment and introduced a new method to approximate the expected cost of any VRPSD with replenishment.

In recent years, research on VRP has become more diverse, with scholars proposing models to optimize different aspects of logistics transportation. For example, Islam *et al.* [19] studied the Clustered Vehicle Routing Problem (CluVRP) and proposed a new hybrid meta-heuristic algorithm combining particle swarm optimization (PSO) and variable neighbourhood search (VNS) to solve the model. Solomon and Desrosiers [20] incorporated the concept of time windows into vehicle routing problems, and Jabali *et al.* [21] proposed a vehicle routing problem with soft and hard time windows. Rodríguez-Martín and Yaman [22] developed a periodic vehicle routing problem with driver consistency, and Yuan *et al.* [23] studied the generalized vehicle routing problem with time windows. Zhao *et al.* [24] considered the departure time and the distance between two customers. They proposed a bi-objective mixed integer linear model to optimize the total transportation cost and time cost.

The Green Vehicle Routing Problem has gained significant attention in recent years, aiming to promote green development by reducing energy consumption and carbon emissions in logistics activities. Scholars have proposed various models to optimize different aspects of logistics transportation while reducing carbon emissions. For example, Demir *et al.* [25] compared and analyzed six models for carbon emissions and energy consumption. Kirschstein and Meiselb [26] designed a comprehensive carbon emission calculation model by considering factors such as vehicle speed, load, and road conditions. Naderipur and Alinaghian [27] studied low carbon VRP with the goal of reducing vehicle energy consumption and carbon emissions. Kwon *et al.* [28] proposed a multi-vehicle VRP model based on carbon emissions to minimize total cost. Suzuki [29] established a VRP model aimed at minimizing energy consumption and carbon emissions. Li *et al.* [30] constructed a VRP model that aimed to minimize the sum of vehicle fixed usage costs, fuel consumption, and carbon emissions costs. Wen *et al.* [31] proposed a multi-site model of vehicle routing to optimize carbon emissions, fuel consumption, vehicle rental, and driver wage costs. They developed an improved adaptive large neighbourhood search (ALNS) algorithm to effectively solve the problem. Guo *et al.* [32] studied the multi-compartment vehicle routing problem considering carbon emissions and optimized the total transportation cost, including carbon emissions, using a three-dimensional ant colony optimization algorithm (TDACO). Li and Li [33] proposed a multi-objective supply chain network optimization model that aimed to optimize network costs, carbon trading costs, and customer satisfaction losses. They developed a new improved NSGA-II algorithm to solve the model. Zhu *et al.* [34] established a CVR for multiple warehouses with the goal of minimizing the carbon emissions of the fleet required to deliver the required goods to customers.

Despite the significant progress made in the field of the Green Vehicle Routing Problem, there is still a need for more effective and comprehensive models that can address the challenges posed by traffic congestion and low-carbon emission reduction in urban logistics network optimization. Therefore, this paper proposes a new model and algorithm to address these gaps in the literature, which has practical significance for the development of low-carbon logistics and the improvement of urban traffic congestion. Urban traffic networks have time-varying characteristics due to factors such as morning and evening traffic peaks, road speed limits, traffic regulations, and external accidents. To address this, some scholars have studied time-dependent vehicle routing problems (TDVRP) under time-varying road networks. For example, Jabbarpour *et al.* [35] established a TDVRP model that aimed to minimize driving time and fuel consumption, and designed different traffic congestion scenarios for experiments. Xiao and Konak [36] proposed that highway transportation companies can reduce their CO_2 emissions through effective vehicle routing and delivery schedules based on traffic congestion in their service areas. Poonthalir and Nadarajan [37] focused on behavior in variable speed environments and its impact on route costs and fuel consumption. They built a TDVRP model that aimed to minimize vehicle travel distance and fuel consumption and designed an improved particle swarm optimization algorithm to solve the problem. Çimen and Soysal [38] considered the vehicle routing problem under time-dependent and random vehicle speeds. The research results showed that incorporating vehicle speed randomness into the model enabled optimization of the final distribution route in terms of travel duration, carbon emissions, and travel costs. Ehmkea *et al.* [39] constructed a TDVRP model that aimed to minimize carbon emissions and solved it using a tabu search algorithm. Sarbijan and Behnamian [40] proposed that in the context of congestion in urban transportation networks, there is a higher requirement for fast, flexible, reliable, and low-cost delivery in urban areas. In real-time collaborative regional vehicle routing problems with flexible time windows, a combination of urban logistics transportation and distribution composed of various vehicles can reduce the number of times to return to physical warehouses, reduce costs, and save time.

Through a comprehensive literature review, scholars have proposed various models to address the challenges posed by time-varying road networks in urban logistics transportation. However, there is still a need for more effective and comprehensive models that can address the challenges posed by traffic congestion, low-carbon emission reduction, and time-varying road networks in urban logistics network optimization. Therefore, this paper proposes a new model and algorithm to address these gaps in the literature, which has practical significance for the development of low-carbon logistics and the improvement of urban traffic congestion. Through a comprehensive literature review, it was found that some research results have been generated on the Vehicle Routing Problem (VRP) under traffic congestion. However, the existing research is limited to constant vehicle speed, and there is a lack of research on VRP under time-varying road networks. Additionally, there is a shortage of research on the impact of traffic congestion and low-carbon emission reduction on urban logistics network optimization.

To address these gaps in the literature, this paper proposes a new multi-objective urban logistics network optimization model that considers traffic congestion to optimize vehicle usage costs, transportation costs, penalty costs, and carbon emission costs. The proposed model takes into account the impact of time-varying road networks and the need for low-carbon logistics. To solve the model, a DPSO algorithm based on the basic principle of PSO is developed.

This research has practical significance for the development of low-carbon logistics and the improvement of urban traffic congestion. By optimizing vehicle routing and reducing carbon emissions, this research can help reduce the negative impact of logistics activities on the environment and promote sustainable development. Moreover, the proposed model and algorithm can provide valuable guidance for logistics companies and transportation departments in urban areas to optimize their logistics network and reduce transportation costs.

3. Optimization model of vehicle routing problem in urban logistics distribution networks

3.1 Problem description

In today's low-carbon economic environment, it is crucial to consider not only conventional costs such as vehicle usage costs, transportation costs, and penalty costs for failing to meet the time window in urban logistics distribution but also the carbon emission costs associated with transportation. Additionally, the impact of urban road traffic conditions on total costs should also be considered. Therefore, this paper proposes a new model for the Vehicle Routing Problem (VRP) in Urban Logistics Distribution Networks (ULDN), which takes into account traffic congestion. The problem can be described as a distribution center providing logistics distribution services to multiple customer points within a specified time. The goal of the proposed model is to comprehensively optimize vehicle usage costs, transportation costs, time penalty costs, and carbon emissions costs in ULDN. The customer location, customer demand, and time window are known, and the urban traffic congestion period and traffic congestion status can be obtained from the transportation department.

The proposed model and algorithm will provide valuable guidance for logistics companies and transportation departments in urban areas to optimize their logistics network and reduce transportation costs while considering the impact of traffic congestion and low-carbon emissions. By reducing carbon emissions and optimizing vehicle routing, this research can help mitigate the negative impact of logistics activities on the environment and promote sustainable development.

3.2 Assumptions

In the proposed model, the following assumptions are made:

(1) Only one distribution center with sufficient supply is considered.
(2) The delivery vehicles are of the same type and depart from the logistics center at different times as needed, returning to the same logistics center afterward.
(3) Customer demand is less than the vehicle capacity, and there is a service time window requirement.
(4) During periods of traffic congestion, vehicles travel at a congested speed, while during non-congested periods, vehicles travel at normal speeds.
(5) The maximum load capacity of each vehicle is fixed, and each customer is served by only one vehicle.
(6) Vehicles generate carbon emissions during driving time and do not generate carbon emissions during the rest of the time.

By considering these assumptions, the proposed model provides a practical and realistic approach for logistics companies and transportation departments to optimize their logistics network while reducing carbon emissions and transportation costs. The proposed model and algorithm can be used to guide logistics companies in making informed decisions on vehicle routing and scheduling, ultimately improving the efficiency and sustainability of urban logistics distribution networks.

3.3 Notations

I Set of customers
I' Set of all nodes in the urban logistics distribution network
J Set of routes
K Set of vehicles
N Set of road section
f_k Fixed departure cost of vehicles k
t_{ijkn} The time consumed by the vehicle k traveling on road segment n in road (i, j)
O Unit time cost of vehicle usage
D Unit human resource cost of vehicles
Q Maximum vehicle load capacity
q_i The demand for customer i;
ct_i Unloading service time at customer i
T_{ik} Waiting time for vehicle k to arrive at customer i in advance
v_{ijkn} The traveling speed of the vehicle k in the section n of the road (i, j)
t_{ijkn} The time consumed by the vehicle k traveling on road segment n in road (i, j)
β_{ijkn} Carbon emission rate of vehicle k in section n of road (i, j) (kg/km)
d_{ijkn} The distance travelled by vehicle k in section n of road (i, j)
δ Unit carbon emission cost (yuan/kg)
g Unit fuel consumption cost (yuan/L);
θ_{ijkn} The fuel consumption rate of the vehicle k in the section n of the road (i, j) (L/km)
RT_{ik} The time when the vehicle k arrives at the customer i
LT_{ik} The time when the vehicle k leaves the customer i
$[E_j, L_j]$ The service time window for customer i
α_{de} The penalty factor when the vehicle arrives early
α_{dl} The penalty factor when the vehicle arrives late
λ_{ij} Traffic congestion coefficient

$$x_{ijk} = \begin{cases} 1 & \text{If vehicle } k \text{ travel in the path } (i,j) \\ 0 & \text{otherwise} \end{cases} \quad i \in I, j \in I, k \in K$$

$$Y_{jk} = \begin{cases} 1 & \text{If vehicle } k \text{ serves consumer } j \\ 0 & \text{otherwise} \end{cases} \quad j \in I, k \in K$$

$$U_{ijkn} = \begin{cases} 1 & \text{If vehicle } k \text{ travel in the path } n \\ 0 & \text{otherwise} \end{cases} \quad \mathrm{i} \in I, \mathrm{j} \in J, \mathrm{k} \in K, n \in N$$

$$r_k = \begin{cases} 1 & \text{If vehicle } k \text{ is used} \\ 0 & \text{otherwise} \end{cases} \quad k \in K$$

3.4 Mathematical model

With the rapid increase in urban cars, traffic congestion has become a common phenomenon in urban areas. The impact of traffic congestion on logistics delivery efficiency and quality makes it necessary to consider the space-time effect in ULDN. The shortest path between two customers should not be based on the shortest spatial distance but rather on the shortest time, which varies due to different road congestion conditions and vehicle speeds during different periods of time. To quantify the degree of road congestion, the paper introduces a traffic congestion coefficient λ_{ij}. The speed on the road during congestion $v_c = v_f/\lambda_{ij}$, where v_f is the vehicle speed when the road is clear. When the vehicle travels on a sub-road section with a sufficiently short distance, the speed can be considered constant based on the actual situation and driving rules.

In the proposed urban logistics network optimization model, the carbon emission cost, vehicle usage cost, transportation cost, and time penalty cost are considered. They are calculated as follows:

(1) Carbon emission cost

Carbon emissions are mainly generated during vehicle transportation, and the carbon emission coefficient is used to calculate the carbon emissions during transportation. The carbon emission coefficient represents the unit carbon emissions of the logistics distribution, which quantifies the total carbon dioxide content in the logistics distribution. The carbon emission cost C_1 is calculated using Eq. 1.

$$C_1 = \sum_{i\in I}\sum_{j\in I}\sum_{k\in K}\sum_{n\in N}\beta_{ijkn}U_{ijkn}\,v_{ijkn}t_{ijkn} \tag{1}$$

(2) Vehicle usage cost

Vehicle usage costs mainly include vehicle departure costs, vehicle rental costs, and labour costs. The vehicle rental cost and labour cost are the product of the total travel time and unit cost. The total travel time is the sum of road travel time, customer service unloading time, and waiting time at the customer point. Therefore, the vehicle management usage cost C_2 is calculated using Eq. 2.

$$C_2 = \sum_{k\in K} r_k\, f_k + \sum_{i\in I}\sum_{j\in I}\sum_{k\in K}\sum_{n\in N} U_{ijkn}\, t_{ijkn}(O+D) + (\sum_{j\in I}\sum_{k\in K} y_{jk}ct_j + \sum_{j\in I}\sum_{k\in K} y_{jk}T_{jk})(O+D) \tag{2}$$

(3) Transportation cost

Transportation cost refers to the fuel consumption cost generated during vehicle transportation. The transportation cost C_3 is calculated using Eq. 3.

$$C_3 = \sum_{i\in I}\sum_{j\in I}\sum_{k\in K}\sum_{n\in N} gU_{ijkn}\,v_{ijkn}t_{ijkn}\theta_{ijkn} \tag{3}$$

(4) Time penalty cost

In the process of urban logistics distribution, scheduling errors, and low delivery efficiency may cause distribution vehicles to miss the specified delivery time. Such delays may negatively impact customers, such as supermarkets and shopping malls, which have their own business hours. Delayed delivery times may increase the cost of the enterprise due to decreased customer satisfaction, and penalties should be imposed accordingly. The penalty cost in the vehicle distribution is calculated using Eq. 4.

$$C_4 = \alpha_{de}\sum_{i\in I}\sum_{k\in K}(E_{ik} - RT_{ik}, 0) + \alpha_{dl}\sum_{i\in I}\sum_{k\in K}(RT_{ik} - L_{ik}, 0) \tag{4}$$

By considering these costs, the proposed model and algorithm can provide valuable guidance for logistics companies and transportation departments in urban areas to optimize their logistics network while reducing transportation costs and carbon emissions. The $E_{ik} - RT_{ik}$ is the waiting time for vehicle k to arrive at customer i in advance, and $RT_{ik} - L_{ik}$ is the waiting time for customer i due to vehicle lateness.

The objective function of the proposed VRP model in ULDN is constructed as follows:

$$min\, Z_1 = \sum_{i\in I}\sum_{j\in I}\sum_{k\in K}\sum_{n\in N}\beta_{ijkn}U_{ijkn}\,v_{ijkn}t_{ijkn} \tag{5}$$

$$min\, Z_2 = \sum_{k\in K} r_k\, f_k + \sum_{i\in I}\sum_{j\in I}\sum_{k\in K}\sum_{n\in N} U_{ijkn}\, t_{ijkn}(O+D) + (\sum_{j\in I}\sum_{k\in K} y_{jk}ct_j + \sum_{j\in I}\sum_{k\in K} y_{jk}T_{jk})(O+D) \tag{6}$$

$$min\, Z_3 = \sum_{i\in I}\sum_{j\in I}\sum_{k\in K}\sum_{n\in N} gU_{ijkn}\, v_{ijkn} t_{ijkn}\theta_{ijkn} \tag{7}$$

$$min\, Z_4 = \alpha_{de}\sum_{i\in I}\sum_{k\in K}(E_{ik} - RT_{ik}, 0) + \alpha_{dl}\sum_{i\in I}\sum_{k\in K}(RT_{ik} - L_{ik}, 0) \tag{8}$$

Subject to

$$\sum_{k\in K} x_{ijk} = 1\,, \forall (i,j) \in J \tag{9}$$

$$\sum_{k\in K} y_{jk} = 1\,, \forall j \in I \tag{10}$$

$$x_{ijk} \geq U_{ijkn}, \forall (i,j) \in J, k \in K, h \in H \tag{11}$$

$$x_{ijk} \leq \textstyle\sum_{n\in N} U_{ijk}\,, \forall (i,j) \in J, k \in K \tag{12}$$

$$\sum_{k\in K}\sum_{n\in N} d_{ijkn} U_{ijkn} = x_{ijk} d_{ijkn}, \forall (i,j) \in J, k \in K, n \in N \tag{13}$$

$$\sum_{j\in I'} x_{0jk} \leq 1\,, \forall k \in K \tag{14}$$

$$d_{ijkn} \leq d_{ij} U_{ijkn}, \forall (i,j) \in J, k \in K, n \in N \tag{15}$$

$$\sum_{j\in J} q_j\, y_{jk} \leq Q, \forall k \in K \tag{16}$$

$$x_{ijk} \in \{0,1\} \tag{17}$$

$$y_{jk} \in \{0,1\} \tag{18}$$

$$U_{ijkn} \in \{0,1\} \tag{19}$$

$$r_k \in \{0,1\} \tag{20}$$

Eq. 9 indicates that only one vehicle is allowed to drive on the selected road. Eq. 10 indicates that each customer can only be served by one vehicle, and all customers must be served. Eqs. 11 and 12 represent the limiting relationship between variables x_{ijk} and U_{ijkn}. Eq. 13 indicates that the vehicle should travel the entire road as long as the road is selected. Eq. 14 indicates that each vehicle is only used once. Eq. 15 represents the restriction relationship between d_{ijkn} and d_{ij}. Eq.16 represents vehicle capacity constraints. Eqs. 17-20 represent variable value constraints.

By considering these constraints and the objective function, the proposed model and algorithm can optimize vehicle usage costs, transportation costs, time penalty costs, and carbon emissions costs in ULDN while considering traffic congestion. This approach can help logistics companies and transportation departments in urban areas to optimize their logistics network and reduce transportation costs and carbon emissions, ultimately promoting sustainable development.

4. Improved Discrete Particle Swarm Optimization

The Particle Swarm Optimization (PSO) is a global optimization algorithm based on swarm intelligence, which was proposed by American scholars Kennedy and Eberhart [41]. The PSO algorithm simulates the social behaviour of animal groups, such as flocks of birds and fish, by following three typical rules: 1) Fly away from the nearest individual to avoid collisions; 2) Fly towards a predetermined goal; 3) Fly to the center of the group. For example, a flock of birds usually determines its flight direction and speed based on its own flight experience, which leads to consistent flock behavior. However, when one bird in the group changes direction and flies to a new habitat, other birds will also be affected and fly to the new habitat, causing the remaining birds to imitate this behaviour until they all fall into the new habitat.

In the PSO algorithm, each possible solution in a population is represented as a particle without volume or mass. All such particles fly at a certain speed in the search space, and their speed is derived from past flight experience. The PSO algorithm enables the entire population to develop towards global optimization through information sharing among particles. It has the ability to search multiple points and can obtain multiple Pareto optimal solutions through one operation. Therefore, the advantages of the PSO algorithm are suitable for solving multi-objective optimization problems in urban logistics networks.

To further improve the PSO algorithm, the paper proposes an improved discrete particle swarm optimization algorithm (DPSO) to solve the VRP model in ULDN. The DPSO algorithm uses multiple populations to process multiple targets and develops a variable neighbourhood search strategy to improve the search ability of particles. The DPSO conducts a randomized deep search to improve the "premature convergence" problem of the PSO algorithm, which improves the quality of understanding.

(1) Coding of the DPSO

To implement the DPSO algorithm for solving the VRP model in ULDN, the solution space of VRP in ULDN is represented by a directed complete graph, denoted as $G = (V, E)$, where each potential solution is a generated subgraph of G. The search space of the entire particle swarm is the arc set E in the complete graph G. The position of each particle is represented by a set consisting of arcs, forming a subset A. The search space of the particle swarm is the edge set of the directed complete graph of urban logistics distribution customer nodes. The position of a particle is a subset of the edge set of a complete graph, and the edges in this subset are connected end-to-end to form a directed Hamilton loop, serving as the distribution path for logistics vehicles.

The speed of a particle is a collection of all nodes, and edges in the speed collection may be selected to build a new location for the particle. Each element in the individual is converted to a number in the floating point interval [0,1]. The velocities of all particles are calculated, and then the element is converted to an integer based on the relative position index.

(2) Particle position update

The DPSO algorithm for solving the VRP model in ULDN uses n candidate solutions to find the optimal solution in the search space. The position vector corresponding to the particle is represented by $X_i = [x_i^0, x_i^1, \ldots, x_i^n]$, where each dimension of X is represented by $x_i^d = [(m, d), (d, k)], m, d \in \{0,1, \ldots, d-1, d+1, n\}, m \neq k$. Each individual in the population represents a feasible solution, and x_i^d is composed of two arcs connected by customer d. n is the total dimension (total number of customers), d represents the current dimension index, m is the predecessor node of customer t (the customer served before customer t), and k is the successor node of customer t (the customer served after customer t). Eq. 21 is used to update the position of particles.

$$v_{i_d}(t+1) = v_{id}(t) + c_1 \times r_1 \times (pb_{id}(t) - x_{id}(t)) + c_2 \times r_2 \times (gb_d(t) - x_{id}(t)) \qquad (21)$$

The velocity of particle i in the d-th dimension at the t-th iteration is represented as $v_{id}(t) \in R$, and $PB_i(t) = (pb_{i1}(t), pb_{i2}(t), \ldots, pb_{id}(t), \ldots, pb_{iD}(t))$ is the position of the individual historical extreme value of particle i up to the t iteration. $GB(t) = (gb_1(t), gb_2(t), \ldots, gb_d(t), \ldots, gb_D(t))$ is the best position experienced by all particles in the population up to the t iteration. Each particle updates its own velocity and location based on individual historical extremum and global optimal value. The t represents the t-th search, $V_i(t)$ represents the speed of the i-th particle, $X_i(t)$ is the current position of the i-th particle, r_1 and r_2 are random numbers between (0,1), and the constants c_1 and c_2 are learning factors, usually taking the same value between 0 and 2.

(3) Particle Speed Update

In the DPSO algorithm for solving the VRP model in ULDN, when the optimization value of particle i changes very little, the speed of the particle is updated according to Eq. 22, where α_i defines the historical optimal value of particle i. For each dimension of each particle in the D-dimensional space, the corresponding dimension of the historical optimal value of one particle is selected from all particles according to a certain probability to learn. Thereby, a learning particle is constructed randomly for each particle i. Each dimension of each particle in the population learns from the optimal solution set with a probability p.

The particle speeds are updated by Eqs. 22 and 23, where *Gbest* is the optimal solution set, and G is the number of candidate solutions in the optimal solution set, which is equal to the number of population sizes.

$$V_i^d = w \times V_i^d + c \times rand_i^d \times \left(Gbest_{f_i(d)}^d - X_i^d\right) \tag{22}$$

$$\rho C_p \left(\frac{\partial T}{\partial t} + u \cdot \nabla T\right) = \nabla(k \nabla T) + \eta \dot{\gamma}^2 \tag{23}$$

(4) Variable neighbourhood search

During the operation of PSO, particle swarm optimization may quickly approach the current optimal position of the population, which can lead to weak global search ability and the "premature convergence" phenomenon. To address this issue, the DPSO algorithm utilizes variable neighbourhood search local search operators, which can expand the local search space and improve the quality of individual search, thereby improving the group optimization ability.

In the DPSO algorithm for solving the VRP model in ULDN, the vehicle route that passes through the least number of customers is selected to determine the individual and global extrema. Then, under certain hypothetical conditions, the customer closest to the current route is selected to be inserted, and it is determined whether at least that customer exists in another vehicle route. If it exists, the location where the total logistics cost of the new route is the smallest is selected, and if it does not exist, it will not be inserted. The optimal solution set is updated after completing the above operation.

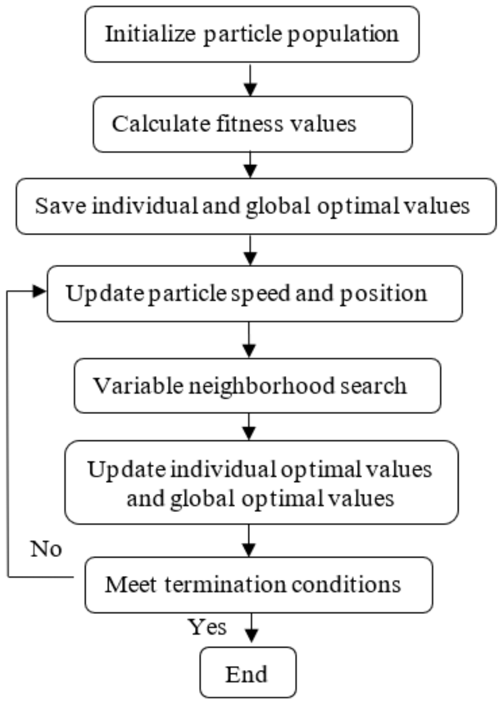


Fig. 1 Algorithm steps of DPSO

The specific steps of variable domain search are as follows:

For $L1(s)$:
Step 1: Select a path R randomly.
Step 2: Select two nodes $N1$ and $N2$ in path R randomly.
Step 3: Exchange $N1$ and $N2$.

For $L2(s)$:
Step 1: Select two paths $R1$ and $R2$ randomly.
Step 2: Select a node $N1$ randomly in path $R1$, and select a node $N2$ randomly in path $R2$.
Step 3: Exchange $N1$ and $N2$.

For $L3(s)$:
Step 1: Select the path R with the lowest bearing capacity.
Step 2: Insert the customers in the distribution path R into other distribution paths.

(5) Algorithm Steps

The DPSO steps is as follows:
Step 1: Initialize the particle population and initialize each particle randomly.
Step 2: Update the speed and position of each particle.
Step 3: Local search on the optimal solution set.
Step 4: Update the optimal solution set. If the stop condition is not satisfied, return to Step 2 to continue optimizing the population.
Step 5: Terminate the entire algorithm and obtain the optimal solution set. The specific steps are shown in Fig. 1.

5. Simulation

Based on the urban logistics distribution data of a logistics enterprise in Shanghai, the DPSO algorithm is applied to optimize the logistics network. The logistics enterprise distributes products to various consumers in Shanghai, and has one logistics distribution center that needs to meet the distribution needs of 50 stores and supermarkets (customers). The number 0 refers to the logistics distribution center, and the numbers 1-50 refer to the consumers.

The initial time is set as 6:00 a.m., and this time is set as 0 o'clock in the model. According to urban traffic laws, the time periods between 7:00 to 9:00 and 17:00 to 19:00 are considered as traffic congestion periods, and the rest of the time periods are considered as normal driving periods. Under a time-varying network, the normal driving speed v_f is 60 km/h, and the speed v_c during congestion is 30 km/h. The vehicle has an unloaded weight of 5000 kg and a capacity of $2t$.

The values of other parameters are shown in Table 1. The parameters of the DPSO algorithm are set as follows: the population size is 50, the number of iterations is 100, $w = 0.7$, $c_1 = c_2 = 1.5$, $r_1 = r_2 = 0.5$.

The simulation experiment using the DPSO algorithm to optimize the logistics network of a logistics enterprise in Shanghai took 180.2 s to execute. Seven vehicles were used for distribution, and the total distribution cost was 5240 yuan. The vehicle usage costs (vehicle startup and rental fees) were 2305 yuan, the transportation cost was 2638 yuan, the carbon emission cost was 103 yuan, and the time penalty cost was 194 yuan. The optimization results are shown in Table 2.

The optimization results indicate that the DPSO algorithm can obtain optimal routes in a relatively short time. Vehicles 3 and 5 completely avoided the congestion period, while vehicles 1, 4, and 6 had two sections in the morning and evening peak congestion periods. Vehicle 2 had two sections in the morning and evening peak congestion period, and vehicle 7 had three sections in the morning and evening peak congestion period. This indicates that the proposed DPSO algorithm can reasonably avoid traffic congestion periods and improve vehicle delivery efficiency.

Table 1 Parameters

Parameters	Value
f_k	130yuan/vehicle
O	6yuan/h
D	10yuan/h
δ	0.5yuan/kg
g	6yuan/L
α_{de}	10yuan/h
α_{dl}	10yuan/h
v_f	60km/h
v_c	30 km/h

Table 2 Optimization results of urban logistics distribution network

Route	Distribution route	Number of routes during congestion periods	Carbon emission cost	Vehicle usage costs	Transport-ation cost	Time penalty cost	Total cost
1	0-4-5-10-12-15-20-21-0	1	11	286	325	26	648
2	0-1-6-7-25-41-8-16-44-0	2	15	355	410	45	825
3	0-32-45-28-0	0	8	167	203	0	378
4	0-14-2-17-19-49-38-26-0	1	14	290	348	23	675
5	0-42-31-23-3-9-18-43-0	0	12	348	322	0	682
6	0-22-11-13-34-47-27-50-36-0	1	19	365	402	21	807
7	0-29-46-24-33-35-30-39-40-48-37-0	3	24	494	628	79	1225

Based on the comprehensive distribution route optimization results, it can be found that due to the constraints of urban congestion and customer service time windows, there are significant differences in the distribution routes. The maximum number of customers served by vehicle 7 is 10, while the number of customers served by vehicle 3 is the lowest, with only 3. This is due to the different time windows at each customer, indicating that logistics enterprises should consider time dependence when planning their routes. Logistics enterprises should plan their routes scientifically based on the actual conditions such as road network conditions and customer time windows.

In addition, the impact of traffic congestion on the optimization results of VRP in ULDN is illustrated in the paper. The traffic congestion coefficient λ_{ij} is set to 1.5, 2.0, 2.5, and 3.0, with corresponding congestion speeds of 40 km/h, 30 km/h, 24 km/h, and 20 km/h, while the normal driving speed v_f remains unchanged at 60 km/h. The optimization results are shown in Table 3.

Table 3 Optimization results for different traffic congestion coefficient

Traffic congestion coefficient	Number of delivery vehicles	Carbon emission cost	Vehicle usage costs	Transportation cost	Time penalty cost	Total cost
1.5	7	89	2260	2554	181	5084
2.0	7	106	2302	2644	199	5251
2.5	8	134	2377	2880	221	5612
3.0	9	165	2456	3012	247	5880

From the perspective of the optimization process of the DPSO algorithm, the optimal evolution iterations corresponding to different traffic congestion coefficients fluctuate slightly between 30 and 45 generations after 100 iterations. Although the change in vehicle usage costs is not significant, the optimal values of transportation costs, carbon emissions costs, and time penalty costs increase as the traffic congestion coefficient increases, leading to an increase in total costs. As the congestion coefficient increases, fuel consumption and carbon emissions increase, indicating that congestion conditions can affect the greenness of logistics delivery routes. The cost of time penalty increases, which indicates that traffic congestion can affect vehicle speed and affect the service time.

Furthermore, the effectiveness of the DPSO algorithm is verified by comparing the optimization results of DPSO with the basic PSO algorithm. The convergence of the two algorithms within 100 iterations is obtained under the same parameter settings. The parameters of the algorithm are: the population size is 50, the number of iterations is 100, $w = 0.7$, $c_1 = c_2 = 1.5$, $r_1 = r_2 = 0.5$. The optimization results of DPSO and PSO are shown in Figure 2.

The optimization results of the DPSO algorithm show that the minimum value of the total cost maintains an overall downward trend with the increase of genetic iterations. Meanwhile, the convergence of the DPSO algorithm is significantly better than that of the PSO algorithm. The DPSO algorithm basically reaches the optimal solution around the 35th generation, while the PSO algorithm converges to the optimal solution in the 80th generation, and its overall optimization cost is greater than that of the DPSO algorithm. This indicates the effectiveness of the DPSO algorithm in optimizing the logistics network in ULDN and improving the efficiency of vehicle delivery.

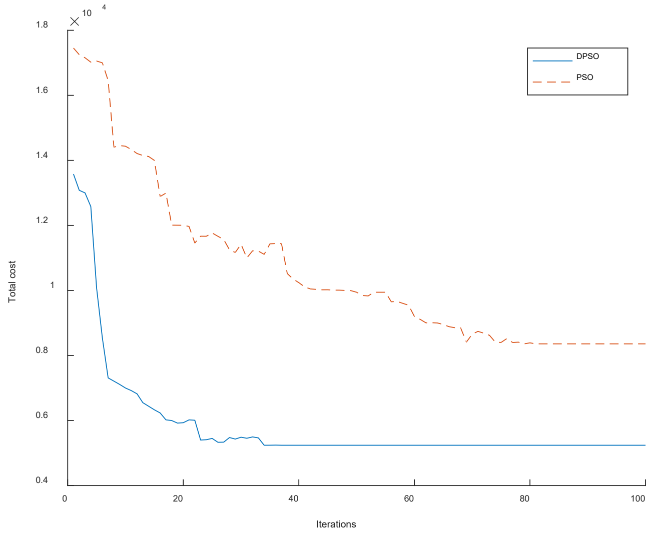


Fig. 2 Comparison of the optimal total cost

6. Conclusions

In this paper, we proposed a model of vehicle routing problem in urban logistics distribution under traffic congestion, taking into account vehicle management costs, transportation costs, carbon emissions costs, and penalty costs comprehensively. To solve this complex multi-objective optimization problem, we developed an improved DPSO algorithm based on the PSO, which uses multiple populations to process multiple targets, and incorporates a variable neighbourhood search strategy to improve the search ability of particles. The randomized deep search was also conducted to improve the "premature convergence" problem of PSO, which improves the quality of optimization results.

The simulation results showed that our proposed model can obtain the lowest cost and optimal delivery route, effectively avoiding traffic congestion, reducing carbon emissions costs and penalty costs, and improving customer satisfaction. Comparative analysis of optimization results between the DPSO and the PSO showed that the proposed DPSO algorithm has better convergence and effectiveness in VRP of urban logistics distribution.

Our proposed model and algorithm provide effective solutions and references for solving practical logistics distribution network optimization problems in urban areas with complex traffic conditions. By optimizing delivery routes, logistics enterprises can reduce transportation costs, carbon emissions, and penalties, while improving customer satisfaction and promoting sustainable development in urban logistics. In summary, our study contributes to the field of urban logistics distribution and provides a valuable reference for future research in this area. The types of vehicles and the dynamic changes in customer demand are not considered in the study. Future research could introduce different types of delivery vehicles and dynamic changes in customer demand in the delivery paths optimization.

Acknowledgment

This research was funded by Youth Fund for Humanities and Social Sciences Research of Chinese Ministry of Education (Grant No.: 18YJCZH192).

References

[1] Wang, J., Lim, M.K., Tseng, M.-L., Yang, Y. (2019). Promoting low carbon agenda in the urban logistics network distribution system, *Journal of Cleaner Production*, Vol. 211, 146-160, doi: 10.1016/j.jclepro.2018.11.123.

[2] Moghaddas, Z., Vaez-Ghasemi, M., Lotfi, F.H. (2021), A novel DEA approach for evaluating sustainable supply chains with undesirable factors, *Economic Computation And Economic Cybernetics Studies And Research*, Vol. 55, No. 2, 177-192, doi: 10.24818/18423264/55.2.21.11.

[3] Švadlenka, L., Bošković, S., Jovčić, S., Simic, V., Kumar, S., Zanne, M. (2023). Third-party logistics provider selection for sustainable last-mile delivery: A case study of e-shop in Belgrade, *Journal of Urban Development and Management*, Vol. 2, No. 1, 1-13, doi: 10.56578/judm020101.

[4] Nikzamir, M., Baradaran, V., Panahi, Y. (2020). Designing a logistic network for hospital waste management: A benders decomposition algorithm, *Environmental Engineering and Management Journal*, Vol. 19, No. 11, 1937-1956, doi: 10.30638/eemj.2020.184.

[5] Sárdi, D.L., Bóna, K. (2021). City logistics analysis of urban areas: An analytic hierarchy process based study, *Journal of System and Management Sciences*, Vol. 11, No. 2, 77-105, doi: 10.33168/JSMS.2021.0206.

[6] Song, Z., Wang, X. (2020). Contract coordination for green logistics capabilities involving the logistics service demand side, *Environmental Engineering and Management Journal*, Vol. 19, No. 9, 1519-1533, doi: 10.30638/eemj.2020.142.

[7] Dantzig, G.B., Ramser, J.H. (1959). The truck dispatching problem, *Management Science*, Vol. 6, No. 1, 80-91, doi: 10.1287/mnsc.6.1.80.

[8] Niu, X.Y., Liu, S.F., Huang, Q.L. (2022). End-of-line delivery vehicle routing optimization based on large-scale neighbourhood search algorithms considering customer-consumer delivery location preferences, *Advances in Production Engineering & Management*, Vol. 17, No. 4, 439-454, doi: 10.14743/apem2022.4.447.

[9] Wang, Y.D., Lu, X.C., Song, Y.M., Feng, Y., Shen, J.R. (2022). Monte Carlo tree search improved genetic algorithm for unmanned vehicle routing problem with path flexibility, *Advances in Production Engineering & Management*, Vol. 17, No. 4, 425-438, doi: 10.14743/apem2022.4.446.

[10] Barros, L., Linfati, R., Escobar, J.W. (2020). An exact approach for the consistent vehicle routing problem (ConVRP), *Advances in Production Engineering & Management*, Vol. 15, No. 3, 255-266, doi: 10.14743/apem2020.3.363.

[11] Wang, S.R., Huang, Q. (2022). A hybrid code genetic algorithm for VRP in public-private emergency collaborations, *International Journal of Simulation Modelling*, Vol. 21, No. 1, 124-135, doi: 10.2507/IJSIMM21-1-595.

[12] Kunnapapdeelert, S., Thawnern C. (2021). Capacitated vehicle routing problem for Thailand's steel industry via saving algorithms, *Journal of System and Management Sciences*, Vol. 11, No. 2, 171-181, doi: 10.33168/JSMS.2021.0211.

[13] Liu, S., Wang, Y. (2023). Dynamic simulation of differential driving cargo transport vehicle, *Journal of Intelligent Management Decision*, Vol. 2, No. 1, 38-45, doi: 10.56578/jimd020105.

[14] Ismail, B., El Enin, M.A., Osama, M., Abdelhaleem, M., Geris, M., Kamel, M., Kassem, S., Fahim, I.S. (2021). A heterogeneous vehicle routing problem with soft time windows for 3PL company's deliveries: A case study, *Journal Européen des Systèmes Automatisés,* Vol. 54, No. 6, 909-914, doi: 10.18280/jesa.540614.

[15] Dabic-Miletic, S. (2023). Autonomous vehicles as an essential component of Industry 4.0 for meeting last-mile logistics requirements, *Journal of Industrial Intelligence*, Vol. 1, No. 1, 55-62, doi: 10.56578/jii010104.

[16] Gulczynski, D., Golden, B., Wasil, E. (2010). The split delivery vehicle routing problem with minimum delivery amounts, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 46, No. 5, 612-626, doi: 10.1016/j.tre.2009.12.007.

[17] Contardo, C., Martinelli, R. (2014). A new exact algorithm for the multi-depot vehicle routing problem under capacity and route length constraints, *Discrete Optimization*, Vol. 12, 129-146, doi: 10.1016/j.disopt.2014.03.001.

[18] Bertazzi, L., Secomandi, N. (2018). Faster rollout search for the vehicle routing problem with stochastic demands and restocking, *European Journal of Operational Research*, Vol. 270, No. 2, 487-497, doi: 10.1016/j.ejor.2018.03.034.

[19] Islam, M.A., Gajpal, Y., ElMekkawy, T.Y. (2021). Hybrid particle swarm optimization algorithm for solving the clustered vehicle routing problem, *Applied Soft Computing*, Vol. 110, Article No. 107655, doi: 10.1016/j.asoc.2021.107655.

[20] Solomon, M.M., Desrosiers, J. (1988). Survey paper – time window constrained routing and scheduling problems, *Transportation Science*, Vol. 22. No. 1, 1-13, doi: 10.1287/trsc.22.1.1.

[21] Jabali, O., Leus, R., Van Woensel, T., De Kok, T. (2015). Self-imposed time windows in vehicle routing problems, *OR Spectrum*, Vol. 37, 331-352, doi: 10.1007/s00291-013-0348-1.

[22] Rodríguez-Martín, I., Yaman, H. (2022). Periodic vehicle routing problem with driver consistency and service time optimization, *Transportation Research Part B: Methodological*, Vol. 166, 468-484, doi: 10.1016/j.trb.2022.11.004.

[23] Yuan, Y., Cattaruzza, D., Ogier, M., Semet, F., Vigo, D. (2021). A column generation based heuristic for the generalized vehicle routing problem with time windows, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 152, Article No. 102391, doi: 10.1016/j.tre.2021.102391.

[24] Zhao, P.X., Luo, W.H., Han, X. (2019). Time-dependent and bi-objective vehicle routing problem with time windows, *Advances in Production Engineering & Management*, Vol. 14, No. 2, 201-212, doi: 10.14743/apem2019.2.322.

[25] Demir, E., Bektaş, T., Laporte, G. (2011). A comparative analysis of several vehicle emission models for road freight transportation, *Transportation Research Part D: Transport and Environment*, Vol. 16, No. 5, 347-357, doi: 10.1016/j.trd.2011.01.011.

[26] Kirschstein, T., Meisel, F. (2015). GHG-emission models for assessing the eco-friendliness of road and rail freight transports, *Transportation Research Part B: Methodological*, Vol. 73, 13-33, doi: 10.1016/j.trb.2014.12.004.

[27] Naderipour, M., Alinaghian, M. (2016). Measurement, evaluation and minimization of CO_2, NO_x, and CO emissions in the open time dependent vehicle routing problem, *Measurement*, Vol. 90, 443-452, doi: 10.1016/j.measurement.2016.04.043.

[28] Kwon, Y.-J., Choi, Y.-J., Lee, D.-H. (2013). Heterogeneous fixed fleet vehicle routing considering carbon emission, *Transportation Research Part D: Transport and Environment*, Vol. 23, 81-89, doi: 10.1016/j.trd.2013.04.001.

[29] Suzuki, Y. (2016). A dual-objective metaheuristic approach to solve practical pollution routing problem, *International Journal of Production Economics*, Vol. 176, 143-153, doi: 10.1016/j.ijpe.2016.03.008.

[30] Li, J., Wang, D., Zhang, J. (2018). Heterogeneous fixed fleet vehicle routing problem based on fuel and carbon emissions, *Journal of Cleaner Production*, Vol. 201, 896-908, doi: 10.1016/j.jclepro.2018.08.075.

[31] Wen, M., Sun, W., Yu, Y., Tang, J., Ikou, K. (2022). An adaptive large neighborhood search for the larger-scale multi depot green vehicle routing problem with time windows, *Journal of Cleaner Production*, Vol. 374, Article No. 133916, doi: 10.1016/j.jclepro.2022.133916.

[32] Guo, N., Qian, B., Na, J., Hu, R., Mao, J.-L. (2022). A three-dimensional ant colony optimization algorithm for multi-compartment vehicle routing problem considering carbon emissions, *Applied Soft Computing*, Vol. 127, Article No. 109326, doi: 10.1016/j.asoc.2022.109326.

[33] Li, D., Li, K. (2023). A multi-objective model for cold chain logistics considering customer satisfaction, *Alexandria Engineering Journal*, Vol. 67, 513-523, doi: 10.1016/j.aej.2022.12.067.

[34] Zhu, X., Zhao, Z., Yan, R. (2020). Low carbon logistics optimization for multi-depot CVRP with backhauls - Model and solution, *Tehnički Vjesnik – Technical Gazette*, Vol. 27, No. 5, 1617-1624, doi: 10.17559/TV-20200809211109.

[35] Jabbarpour, M.R., Noor, R.M., Khokhar, R.H. (2015). Green vehicle traffic routing system using ant-based algorithm, *Journal of Network and Computer Applications*, Vol. 58, 294-308, doi: 10.1016/j.jnca.2015.08.003.

[36] Xiao, Y., Konak, A. (2017). A genetic algorithm with exact dynamic programming for the green vehicle routing & scheduling problem, *Journal of Cleaner Production*, Vol. 167, 1450-1463, doi: 10.1016/j.jclepro.2016.11.115.

[37] Poonthalir, G., Nadarajan, R. (2018). A fuel efficient green vehicle routing problem with varying speed constraint (F-GVRP), *Expert Systems with Applications*, Vol. 100, 131-144, doi: 10.1016/j.eswa.2018.01.052.

[38] Çimen, M., Soysal, M. (2017). Time-dependent green vehicle routing problem with stochastic vehicle speeds: An approximate dynamic programming algorithm, *Transportation Research Part D: Transport and Environment*, Vol. 54, 82-98, doi: 10.1016/j.trd.2017.04.016.

[39] Ehmke, J.F., Campbell, A.M., Thomas, B.W. (2016). Vehicle routing to minimize time-dependent emissions in urban areas, *European Journal of Operational Research*, Vol. 251, No. 2, 478-494, doi: 10.1016/j.ejor.2015.11.034.

[40] Salehi Sarbijan, M., Behnamian, J. (2022). Real-time collaborative feeder vehicle routing problem with flexible time windows, *Swarm and Evolutionary Computation*, Vol. 75, Article No. 101201, doi: 10.1016/j.swevo.2022.101201.

[41] Kennedy, J., Eberhart, R.C. (1997). A discrete binary version of the particle swarm algorithm, In: *Proceedings of 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation,* Orlando, Florida, USA, 4104-4108, doi: 10.1109/ICSMC.1997.637339.

APEM journal

Advances in Production Engineering & Management
Volume 18 | Number 2 | June 2023 | pp 225–236
https://doi.org/10.14743/apem2023.2.469

ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Blockchain-based tripartite evolutionary game study of manufacturing capacity sharing

Wang, T.Y.[a], **Zhang, H.**[a,*]

[a]Jiangsu University of Science and Technology, School of Economics and Management, Zhenjiang, P.R. China

A B S T R A C T

In the context of the new round of manufacturing innovation, the sharing economy drives the transformation of manufacturing industry to accelerate the integration and development. However, there are some problems in the process of manufacturing capacity sharing, such as information privacy and security, and difficulty in tracing the sharing process, etc. The application of blockchain technology can effectively solve these problems. To explore the capacity sharing behaviour of manufacturing enterprises from the perspective of blockchain, the article combines evolutionary game theory and constructs a tripartite game model of manufacturing capacity sharing. The replication dynamics and evolutionary stability of the model are analysed using evolutionary game theory, and numerical simulations are carried out using MATLAB software to analyse the impact of parameter changes on the evolutionary outcome. The research results show that the incentive and penalty coefficients under blockchain technology have a facilitating effect on enterprises to carry out sharing, and the enhancement of reputation gain coefficient and loss can promote positive services on the platform.

A R T I C L E   I N F O

Keywords:
Blockchain;
Manufacturing;
Capacity sharing;
Tripartite evolutionary game;
Simulation;
MATLAB

**Corresponding author:*
haozh168@aliyun.com
(Zhang, H.)

Article history:
Received 17 November 2022
Revised 15 June 2023
Accepted 29 June 2023

1. Introduction

With the development of the Internet of Things, big data, artificial intelligence and a new round of manufacturing reform, the sharing economy has driven the transformation of the manufacturing industry to accelerate its development and provided a transformation direction for the manufacturing industry [1-5]. Manufacturing capacity sharing is an important element in deepening the integration and development of manufacturing and the Internet, with broad development prospects, and it is particularly important to realize the sustainable development of manufacturing capacity sharing.

The imbalance between supply and demand in manufacturing capacity is a common problem in the market, and capacity sharing can alleviate the mismatch between supply and demand [6]. Capacity sharing can only be achieved by "the platform, the companies that demand manufacturing capacity and the companies that own the manufacturing capacity". The strategic choices of platforms and enterprises play an important role in the capacity utilisation of the manufacturing industry and the development of the economy [7, 8]. However, in the process of manufacturing capacity sharing operation, there will be some problems, such as information privacy and security, the sharing process is difficult to trace and transaction supervision difficulties. Blockchain is the underlying technology for many digital cryptocurrencies, and its features such as decentralisation, open ledger, hashing algorithm, and asymmetric encryption [9, 10] can avoid the risk of information leakage during transactions and make them more secure. These features of blockchain coincide with the demand problems that exist in manufacturing capacity sharing, which can improve the efficiency of the platform and effectively solve the problems that exist in the sharing platform [11, 12].

In previous research, some scholars have applied games to production control on the shop floor to effectively deal with production control problems involving multiple production lines or production goals [13]. Xiao M and Tian Z Y proposed a framework for cloud manufacturing capacity sharing based on a cooperative game algorithm and using MATLAB to analyse the evolutionary outcome [14]. Some scholars have applied blockchain technology to agricultural supply chains, supply chain management, and the financial industry [15-18], but in the manufacturing industry it is mostly applied to manufacturing supply chains, industry 4.0 sustainability, and SCQM [19-21], and few scholars have applied blockchain technology to manufacturing capacity sharing.

Accordingly, the article will combine the characteristics of blockchain to construct a blockchain-based manufacturing capacity sharing model and use evolutionary games to study the capacity sharing behaviour of the manufacturing industry in the blockchain environment.

2. Blockchain-based tripartite evolutionary game analysis of manufacturing capacity sharing

2.1 Main principles of blockchain technology

(1) Distributed ledger technology. A data storage technology is a decentralised distributed database. The data in distributed ledger technology is shared, replicated, and synchronized among the nodes, and it records the transactions between the nodes without the involvement of third parties. Each piece of data in a distributed ledger is signed with a complete and unique timestamp and digital cryptography and cannot be tampered with.
(2) Asymmetric encryption. Asymmetric encryption refers to the encryption of data using different ciphers, i.e. a public key and a private key. Blockchain uses asymmetric encryption algorithms to improve the reliability of data. The public key is a cipher that everyone knows and can be used to encrypt data information, while the private key is a cipher used to decrypt data information and only the recipient of the data information has the private key.
(3) Smart contracts. A smart contract is essentially a program whose content is infinitely scalable and is fully distributed. If both or more parties to the contract meet the triggering conditions, the contract will be automatically triggered and irrevocable, and the execution of the contract will be published to the whole network, with all information immutable, i.e. the transaction is traceable, transparent and irreversible.

2.2 Problem analysis

As shown in the left diagram of Fig. 1., without blockchain technology, enterprises need to publish their information to the platform, and then the platform will match the enterprises for transactions, and the symmetry and openness of information between the two enterprises in the transaction matching process are relatively low, and there is a risk of information privacy being leaked. In the right figure, when blockchain technology is introduced, the decentralized feature of blockchain can reduce the role of platform domination, and the information symmetry and openness between the two sides of enterprises in the transaction matching process is higher, which avoids privacy leakage and improves sharing efficiency [22, 23].

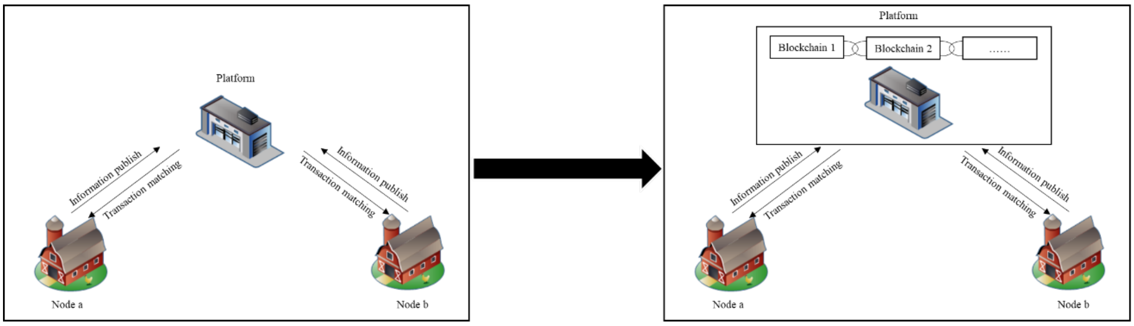


Fig. 1 Before and after introducing blockchain technology

2.3 Coupling analysis of blockchain manufacturing capacity sharing

Coupling analysis refers to the process of considering the interaction or cross-influence of multiple disciplines in a finite analysis. The article constructs a blockchain-based manufacturing capacity sharing model as a fusion innovation for the development of traditional manufacturing capacity sharing platform model, and conducts a coupling analysis between blockchain and manufacturing capacity sharing from three aspects: resource utilization, data trust, and benefit optimization, as shown in Fig. 2.

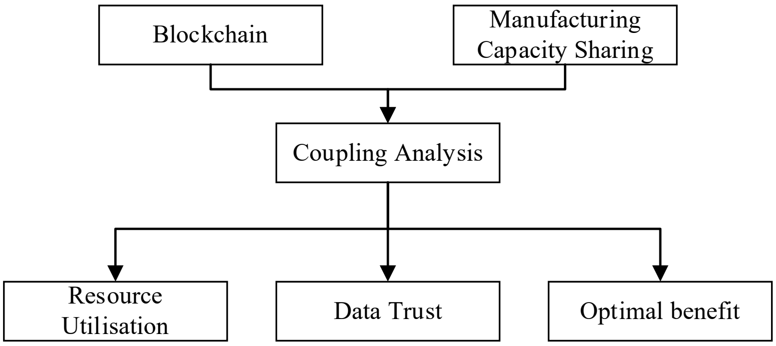


Fig. 2 Analysis of the coupling of blockchain and manufacturing capacity sharing

(1) The coupling of blockchain and manufacturing capacity sharing resource utilization. The capacity provider uploads the redundant capacity information to the platform, and the capacity demander can seek the capacity they need from the platform to maximize resource utilization. The platform data under blockchain technology are all public, and all nodes can query information through the public interface.
(2) The coupling of blockchain and manufacturing capacity sharing data trust. Blockchain technology uses consensus-based specifications and protocols (e.g., open and transparent algorithms) to allow all nodes of the system to exchange data freely and securely in a trusted environment, shifting from trust in the "enterprise" to trust in the "technology".
(3) Optimization of the coupling benefits of blockchain and manufacturing capacity sharing. The purpose of manufacturing capacity sharing is to maximize the utilization of equipment, machines, etc. by integrating and allocating unused capacity. The demanders of capacity can use the capacity, while the providers of capacity can gain revenue and the platform can gain reputation through good service to achieve the best overall benefits for all three parties. One of the features of blockchain is decentralization, where each node is independent and nodes interact with each other without paying additional fees, thus increasing the overall benefits.

2.4 Model assumptions and construction

The article uses evolutionary game theory to investigate the capacity sharing behaviour of manufacturing firms from a blockchain perspective, with each assumption as follows.

Assumption 1. In the game of manufacturing capacity sharing behaviour, there are three nodes of interest, ownership of the capacity, demand for the capacity and the platform of the participant.

Assumption 2. All three nodes have two possible strategies, and will continuously adjust their strategies according to the gains and losses they obtain. Node a has a strategy space of $a = (a_1, a_2)$ = (shared, unshared) and chooses a_1 with probability x and a_2 with probability $(1 - x)$. Node b has a strategy space of $b = (b_1, b_2)$ = (shared, unshared) and chooses b_1 with probability y and b_2 with probability $(1 - y)$. Platform o has a strategy space of $o = (o_1, o_2)$ = (active service , negative service), with probability z of choosing o_1 and probability $(1 - z)$ of choosing o_2 where, $0 \le x, y, z \le 1$.

Assumption 3. When only one party shares, the sharing party pays the corresponding cost C_j, but the digitization level of its enterprise is also improved; the digitization level coefficient d, which brings the benefit e_j, then the digitization benefit obtained by the enterprise is de_j [24]. The manufacturing alliance nodes under blockchain technology have an incentive coefficient m for the sharing party and a penalty coefficient n for the non-sharing party.

Assumption 4. When both nodes share, they receive additional synergy gains under blockchain technology with a synergy gain coefficient θ; the number of shared enterprises is Q_j, and quantifying the capacity sharing level of enterprises [25], the sharing level is G_j. When both nodes do not share, the base gain of enterprises is R_j.

Assumption 5. The horizontal revenue coefficient of the platform under blockchain technology is R_hh when the platform is actively serving, and R_l when it is negatively serving, and $R_h > R_l$. The fixed cost of the platform is C_h when it is actively serving, and C_l when it is negatively serving, and $C_h > C_l$. The platform will bring good reputation revenue for itself when it is actively serving, and the revenue coefficient The probability of matching AB when the firm shares both AB when serving positively is K_hh, the probability of matching AB when serving negatively is K_l, and $K_h > K_l$.

Constructing a revenue matrix is in Table 1:

Table 1 Game payoff matrix

	Node a shares x		Node a does not share $1 - x$	
	Node b shares y	Node b does not share $1 - yz$	Node b shares y	Node b does not share $1 - y$
Node o active service z	$(R_h + F_h)(Q_a + Q_b) - C_h$ $de_a - C_a + mG_a + \theta K_h G_a$ $de_b - C_b + mG_b + \theta K_h G_b$	$(R_h + F_h)Q_a - C_h$ $de_a - C_a + mG_a$ $R_b - nG_a$	$(R_h + F_h)Q_b - C_h$ $R_a - nG_b$ $de_b - C_b + mG_b$	$-C_h$ R_a R_b
Node o negative service 1-z	$(R_l - F_l)(Q_a + Q_b) - C_l$ $de_a - C_a + mG_a + \theta K_l G_a$ $de_b - C_b + mG_b + \theta K_l G_b$	$(R_l - F_l)Q_a - C_l$ $de_a - C_a + mG_a$ $R_b - nG_a$	$(R_l - F_l)Q_b - C_l$ $R_a - nG_b$ $de_b - C_b + mG_b$	$-C_l$ R_a R_b

3. Model analysis

3.1 Analysis of replication dynamics and evolutionary stabilization strategies for node *a*

Based on the gain matrix in the table above, the gains when node a is shared and when it is not shared can be derived as:

$$\begin{aligned} U_{a_1} = yz(de_a - C_a + mG_a + \theta K_h G_a) + y(1-z)(de_a - C_a + mG_a + \theta K_l G_a) \\ + z(1-y)(de_a - C_a + mG_a) + (1-y)(1-z)(de_a - C_a + mG_a) \end{aligned} \tag{1}$$

$$U_{a_2} = yz(R_a - nG_b) + y(1-z)(R_a - nG_b) + z(1-y)R_a + (1-y)(1-z)R_a \tag{2}$$

Therefore, the average expected return of node a and the replication dynamic equation are:

$$U_a = xU_{a_1} + (1-x)U_{a_2} \tag{3}$$

$$\begin{aligned} F(x) = \frac{dx}{dt} = x\left(U_a - U_{a_1}\right) = x(1-x)\left(U_{a_1} - U_{a_2}\right) \\ = x(1-x)[yz\theta G_a(K_h - K_l) + y\theta K_l G_a + ynG_b + de_a - C_a + mG_a - R_a] \end{aligned} \tag{4}$$

The evolutionary strategy of node a must satisfy $F(x) = 0$ and $F'(x) < 0$. For $F(x) = 0$, the solution is $x^* = 0$, $x^* = 1$ and $z^* = \frac{C_a - de_a + R_a - mG_a - y\theta K_l G_a - ynG_b}{y\theta G_a(K_h - K_l)}$. $F'(x) = (1-2x)[yz\theta G_a(K_h -$

$K_l) + y\theta K_l G_a + ynG_b + de_a - C_a + mG_a - R_a]$, make $S(y) = yz\theta G_a(K_h - K_l) + y\theta K_l G_a + ynG_b + de_a - C_a + mG_a - R_a$. Since $S'(y) > 0$, $S(y)$ is an increasing function with respect to y. Therefore, when $z = z^*$, $S(y) = 0$, at which point $F'(x) = 0$, no stable strategy can be determined; when $z < z^*$, $S(y) < 0$, at which point $x = 0$ satisfies the stability condition; when $z > z^*$, $S(y) > 0$, at which point $x = 1$ is ESS, satisfying the stability condition, and the evolutionary phase diagram is shown in Fig. 3.

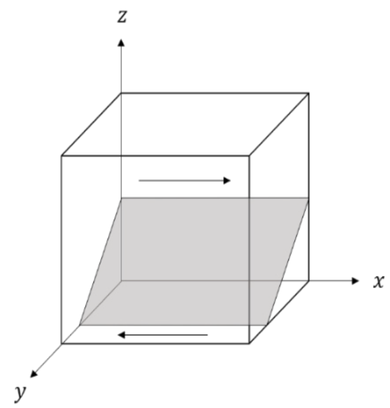


Fig. 3 Digital phase diagrams of node a

3.2 Analysis of replication dynamics and evolutionary stabilization strategies for node *b*

The gains when node b is shared versus not shared are:

$$\begin{aligned} U_{b_1} &= xz(de_b - C_b + mG_b + \theta K_h G_b) + x(1-z)(de_b - C_b + mG_b + \theta K_l G_b) \\ &\quad + z(1-x)(de_b - C_b + mG_b) + (1-x)(1-z)(de_b - C_b + mG_b) \end{aligned} \tag{5}$$

$$U_{b_2} = xz(R_b - nG_a) + x(1-z)(R_b - nG_a) + z(1-x)R_b + (1-x)(1-z)R_b \tag{6}$$

Therefore, the average expected return of node b and the replication dynamic equation are:

$$U_b = yU_{b_1} + (1-y)U_{b_2} \tag{7}$$

$$\begin{aligned} F(y) &= \frac{dy}{dt} = y\left(U_b - U_{b_1}\right) = y(1-y)\left(U_{b_1} - U_{b_2}\right) \\ &= y(1-y)[xz\theta G_b(K_h - K_l) + x\theta K_l G_b + xnG_a + de_b - C_b + mG_b - R_b] \end{aligned} \tag{8}$$

The evolutionary strategy of node b must satisfy $F(y) = 0$ and $F'(y) < 0$. For $F(y) = 0$, the solution is$y^* = 0$, $y^* = 1$ and $z^* = \frac{C_b - de_b + R_b - mG_b - x\theta K_l G_b - xnG_a}{x\theta G_b(K_h - K_l)}$. $F'(y) = (1-2y)[z\theta G_b(K_h - K_l) + x\theta K_l G_b + xnG_a + de_b - C_b + mG_b - R_b]$, make $S(x) = z\theta G_b(K_h - K_l) + x\theta K_l G_b + xnG_a + de_b - C_b + mG_b - R_b$. Since $S'(x) > 0$, $S(x)$ is an increasing function with respect to x. Therefore, when $z = z^*$, $S(x) = 0$, at which point $F'(y) = 0$, no stable strategy can be determined; when $z < z^*$, $S(x) < 0$, at which point $y = 0$ satisfies the stability condition; when $z > z^*$, $S(x) > 0$, at which point $y = 1$ is ESS, satisfying the stability condition, and the evolutionary phase diagram is shown in Fig. 4.

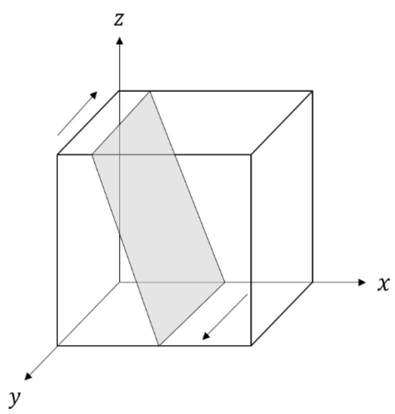


Fig. 4 Digital phase diagrams of node b

3.3 Analysis of replication dynamics and evolutionary stabilization strategies for node *c*

The gains when node o is shared versus not shared are:

$$\begin{aligned} U_{o_1} &= xy[(R_h+F_h)(Q_a+Q_b)-C_h]+x(1-y)[(R_h+F_h)Q_a-C_h] \\ &\quad + y(1-x)[(R_h+F_h)Q_b-C_h]+(1-x)(1-y)(-C_h) \end{aligned} \tag{9}$$

$$\begin{aligned} U_{o_2} = xy[(R_l-F_l)(Q_a+Q_b)-C_l]+x(1-y)[(R_l-F_l)Q_a-C_l]+y(1-x)[(R_l-F_l)Q_b \\ -C_l]+(1-x)(1-y)(-C_l) \end{aligned} \tag{10}$$

Therefore, the average expected return of node o and the replication dynamic equation are:

$$U_o = zU_{o_1}+(1-z)U_{o_2} \tag{11}$$

$$\begin{aligned} F(z) = \frac{dz}{dt} = z\left(U_o-U_{o_1}\right) = z(1-z)\left(U_{o_1}-U_{o_2}\right) \\ = z(1-z)[xQ_a(R_h+F_h-R_l+F_l)+yQ_b(R_h+F_h-R_l+F_l)-C_h+C_l] \end{aligned} \tag{12}$$

The evolutionary strategy of node o must satisfy $F(z)=0$ and $F'(z)<0$. For $F(z)=0$, the solution is $z^*=0$, $z^*=1$ and $x^*=\frac{C_h-C_l-yQ_b(R_h+F_h-R_l+F_l)}{Q_a(R_h+F_h-R_l+F_l)}$. $F'(z)=(1-2z)[xQ_a(R_h+F_h-R_l+F_l)+yQ_b(R_h+F_h-R_l+F_l)-C_h+C_l]$, make $S(y)=xQ_a(R_h+F_h-R_l+F_l)+yQ_b(R_h+F_h-R_l+F_l)-C_h+C_l$. Since $S'(y)>0$, $S(y)$ is an increasing function with respect to y. Therefore, when $x=x^*$, $S(y)=0$, at which point $F'(z)=0$, no stable strategy can be determined; when $x<x^*$, $S(y)<0$, at which point $z=0$ satisfies the stability condition; when $x>x^*$, $S(y)>0$, at which point $z=1$ is ESS, satisfying the stability condition, and the evolutionary phase diagram is shown in Fig. 5.

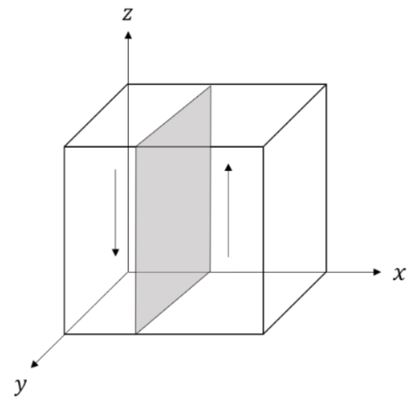


Fig. 5 Digital phase diagrams of node *o*

4. Stability analysis of the equilibrium point of a tripartite evolutionary game system

Letting the replicated dynamic equations equal zero, it is known that there are the following system equilibria E1(0,0,0), E2(0,0,1), E3(0,1,0), E4(0,1,1), E5(1,0,0), E6(1,0,1), E7(1,1,0) and E8(1,1,1), where the Jacobian matrix of the tripartite evolutionary game system is

$$J = \begin{bmatrix} J_1 & J_2 & J_3 \\ J_4 & J_5 & J_6 \\ J_7 & J_8 & J_9 \end{bmatrix} = \begin{bmatrix} \frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y} & \frac{\partial F(x)}{\partial z} \\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y} & \frac{\partial F(y)}{\partial z} \\ \frac{\partial F(z)}{\partial x} & \frac{\partial F(z)}{\partial y} & \frac{\partial F(z)}{\partial z} \end{bmatrix}$$

$$J_1 = (1-2x)[yz\theta G_a(K_h-K_l)+y\theta K_l G_a+ynG_b+de_a-C_a+mG_a-R_a]$$

$$J_2 = x(1-x)[z\theta G_a(K_h-K_l)+\theta K_l G_a+nG_b]$$

$$J_3 = x(1-x)[y\theta G_a(K_h-K_l)]$$

$J_4 = y(1-y)[z\theta G_b(K_h - K_l) + \theta K_l G_b + nG_a]$
$J_5 = (1-2y)[xz\theta G_b(K_h - K_l) + x\theta K_l G_b + xnG_a + de_b - C_b + mG_b - R_b]$
$J_6 = y(1-y)[x\theta G_b(K_h - K_l)]$
$J_7 = z(1-z)[Q_a(R_h + F_h - R_l + F_l)]$
$J_8 = z(1-z)[R_b(R_h + F_h - R_l + F_l)]$
$J_9 = (1-2z)[xQ_a(R_h + F_h - R_l + F_l) + yQ_b(R_h + F_h - R_l + F_l) - C_h + C_l]$

Using Lyapunov first method, the stability of each equilibrium point is analysed as shown [26].

Table 2 Equilibrium point analysis

Equilibrium point	λ1, λ2, λ3	Symbol	Equilibrium results
E1(0,0,0)	$de_a - C_a + mG_a - R_a$, $de_b - C_b + mG_b - R_b$, $-C_h + C_l$	(-,-,-)	ESS
E2(0,0,1)	$de_a - C_a + mG_a - R_a$, $de_b - C_b + mG_b - R_b$, $(-1)(-C_h + C_l)$	(-,-,+)	Unstable
E3(0,1,0)	$\theta K_l G_a + nG_b + de_a - C_a + mG_a - R_a$, $(-1)(de_b - C_b + mG_b - R_b)$, $Q_b(R_h + F_h - R_l + F_l) - C_h + C_l$	(?,+,?)	Unstable
E4(0,1,1)	$\theta G_a(K_h - K_l) + \theta K_l G_a + nG_b + de_a - C_a + mG_a - R_a$, $(-1)(de_b - C_b + mG_b - R_b)$, $(-1)[Q_b(R_h + F_h - R_l + F_l) - C_h + C_l]$	(?,+,?)	Unstable
E5(1,0,0)	$(-1)(de_a - C_a + mG_a - R_a)$, $\theta K_l G_b + nG_a + de_b - C_b + mG_b - R_b$, $Q_a(R_h + F_h - R_l + F_l) - C_h + C_l$	(+,?,?)	Unstable
E6(1,0,1)	$(-1)(de_a - C_a + mG_a - R_a)$, $\theta G_b(K_h - K_l) + \theta K_l G_b + nG_a + de_b - C_b + mG_b - R_b$, $(-1)[Q_a(R_h + F_h - R_l + F_l) - C_h + C_l]$	(+,?,?)	Unstable
E7(1,1,0)	$(-1)(\theta K_l G_a + nG_b + de_a - C_a + mG_a - R_a)$, $(-1)(\theta K_l G_b + nG_a + de_b - C_b + mG_b - R_b)$, $Q_a(R_h + F_h - R_l + F_l) + Q_b(R_h + F_h - R_l + F_l) - C_h + C_l$	(?,?,?)	Unstable
E8(1,1,1)	$(-1)(\theta G_a K_h + nG_b + de_a - C_a + mG_a - R_a)$, $(-1)(\theta G_b K_h + nG_a + de_b - C_b + mG_b - R_b)$, $(-1)[Q_a(R_h + F_h - R_l + F_l) + Q_b(R_h + F_h - R_l + F_l) - C_h + C_l]$	(?,?,?)	Unstable

(1) E_1(0,0,0): $\lambda_1 = de_a - C_a + mG_a - R_a$, $\lambda_2 = de_b - C_b + mG_b - R_b$, $\lambda_3 = -C_h + C_l$.

Scenario 1: $de_a - C_a + mG_a < R_a$, $de_b - C_b + mG_b < R_b$, $C_l < C_h$. Then for nodes a and b, the benefit when only one node shares is less than the benefit when neither node shares; when neither node shares, the cost of negative platform service is less than the cost of positive platform service. At this point λ_1, λ_2 and λ_3 are all less than zero, and the point is the ESS stability point.

(2) E_7(1,1,0): $\lambda_1 = (-1)(\theta K_l G_a + nG_b + de_a - C_a + mG_a - R_a)$, $\lambda_2 = (-1)(\theta K_l G_b + nG_a + de_b - C_b + mG_b - R_b)$, $\lambda_3 = Q_a(R_h + F_h - R_l + F_l) + Q_b(R_h + F_h - R_l + F_l) - C_h + C_l$, at this point the values of λ1, λ2 and λ3 are all uncertain and are discussed by case.

Scenario 2: If $\theta K_l G_a + nG_b + de_a - C_a > R_a - mG_a$, $\theta K_l G_b + nG_a + de_b - C_b > R_b - mG_b$, $(Q_a + Q_b)(R_h + F_h) - C_h < (Q_a + Q_b)(R_l - F_l) - C_l$, then in the case of platform negative service, the gain when both nodes share is greater than the gain when only one side shares; when both sides share, the gain of platform positive service is less than the gain of platform negative service. At this point λ_1, λ_2 and λ_3 are all less than zero, and the point is the ESS point.

Scenario 3: If $\theta K_l G_a + nG_b + de_a - C_a > R_a - mG_a$, $\theta K_l G_b + nG_a + de_b - C_b > R_b - mG_b$, $(Q_a + Q_b)(R_h + F_h) - C_h > (Q_a + Q_b)(R_l - F_l) - C_l$, then in the case of platform negative service, the gain when both nodes share is greater than the gain when only one side shares; when both sides share, the gain from platform positive service is greater than the gain from platform negative service. At this point there is an eigenvalue greater than zero in λ_1, λ_2 and λ_3, which is not an ESS point.

(3) E_8(1,1,1): $\lambda_1 = (-1)(\theta G_a K_h + nG_b + de_a - C_a + mG_a - R_a)$, $\lambda_2 = (-1)(\theta G_b K_h + nG_a + de_b - C_b + mG_b - R_b)$, $\lambda_3 = (-1)[Q_a(R_h + F_h - R_l + F_l) + Q_b(R_h + F_h - R_l + F_l) - C_h + C_l]$, at this point the values of λ1, λ2 and λ3 are all uncertain and are discussed by case.

Scenario 4: If $\theta G_a K_h + nG_b + de_a - C_a > R_a - mG_a$, $\theta G_b K_h + nG_a + de_b - C_b > R_b - mG_b$, $(Q_a + Q_b)(R_h + F_h) - C_h > (Q_a + Q_b)(R_l - F_l) - C_l$, then, in the case of positive platform service, the gain when both nodes share is greater than the gain when only one side shares; when both sides share, the gain from positive platform service is greater than the gain from negative service, when λ_1, λ_2 and λ_3, are all less than zero, and the point is the ESS point.

Scenario 5: If $\theta G_a K_h + nG_b + de_a - C_a < R_a - mG_a$, $\theta G_b K_h + nG_a + de_b - C_b < R_b - mG_b$, $(Q_a + Q_b)(R_h + F_h) - C_h < (Q_a + Q_b)(R_l - F_l) - C_l$, then, in the case of platform positive service, the gain when both nodes share is less than the gain when only either one shares; when both share, the gain from platform positive service is less than the gain from negative service. At this point there is an eigenvalue greater than zero in λ_1, λ_2 and λ_3, which is not an ESS point.

5. MATLAB simulation analysis

To verify the validity of the evolutionary stability analysis, the article incorporates a three-way evolutionary game model, gives the model initial values and uses MATLAB for simulation. The initial values are given in the following table.

The initial values were evolved 50 times over time from different combinations of policies, respectively. As can be seen in Fig. 6, regardless of the initial probability of policy selection for the tripartite nodes, the evolutionary result tends to be (1,1,1), with the corresponding evolutionary policy being (shared, shared, active service), at which point the evolutionary result satisfies Scenario four.

Table 3 Initial values of variables

Variables	Initial value	Variables	Initial value
C_a	5	R_b	10
C_b	5	R_h	0.7
d	0.5	R_l	0.5
e_a	6	C_h	6
e_b	6	C_l	3
m	0.6	F_h	0.6
n	0.2	F_l	0.2
θ	0.4	K_h	0.8
G_a	20	K_l	0.5
G_b	20	Q_a	15
R_a	10	Q_b	15

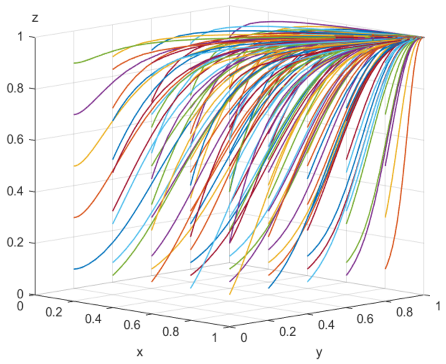


Fig. 6 Evolutionary results for scenario four

Change the initial values so that $R_a = 20$ and $R_b = 20$, and evolve the values 50 times over time from different strategy combinations respectively. As can be seen from Fig. 7, there is only one evolutionary stable strategy combination for the evolving system at this point, and the result satisfies case one, i.e. it eventually converges to (0,0,0) and the corresponding evolutionary strategy is (no share, no share, negative service).

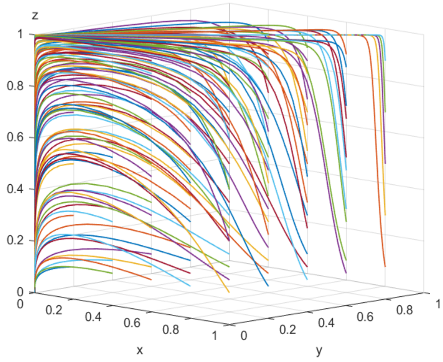


Fig. 7 Evolutionary results for scenario one

Change the value of C_h to satisfy case two. The platform is reluctant to provide positive services because the cost of positive services is too high, at which point the system evolves to a stable point (shared, shared, negative services). With blockchain technology, a reasonably set incentive and penalty for the platform to fulfil its duty to provide good service when both parties share, promoting the sharing of manufacturing capacity and avoiding negative service, as shown in Fig 8.

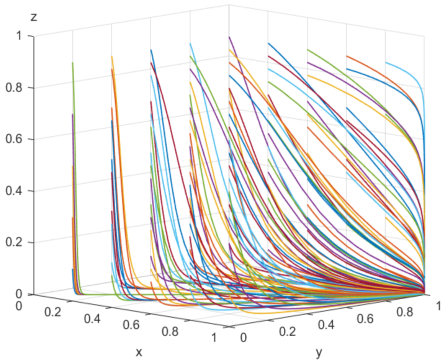


Fig. 8 Evolutionary results for scenario two

To explore the effects of changes in the initial probabilities and other parameters on the evolutionary results, P was adjusted to 0.25, 0.45 and 0.65; m was adjusted to 0.2, 0.6 and 1; n was adjusted to 0, 0.3 and 0.6; d was adjusted to 0, 0.5 and 1; F_h was adjusted to 0.2, 0.6 and 1; and F_l was adjusted to 0, 0.2, 0.4. 0.4. Observe the dynamic course of the evolutionary results over time, as shown in Figs. 9 to 14.

We can see that: as P increases, the probability that the tripartite nodes tend to share, share and actively serve increases, and the speed of evolution gradually increases. When the incentive coefficient m of blockchain technology for nodes is too low, both nodes are unwilling to share, then the platform will pay more costs for positive service than negative service, and eventually the platform will gradually choose to provide negative service; when m increases, the probability of nodes a and b sharing and platform positive service increases, and the convergence speed gradually accelerates to (1,1,1). As the penalty coefficient n increases, the evolution of both nodes choosing to share and platform positive services converge to 1 and the evolution speed gradually accelerates. Penalties under blockchain technology have a positive effect on promoting inter-node behaviour and can effectively facilitate capacity sharing between enterprises.

When the digital revenue d is too low, it will lead to the revenue when nodes share is less than the revenue when they do not share, and the nodes gradually evolve to (0,0,0). A reasonable d will promote the behaviour of sharing among enterprise nodes and active service of platform nodes, and the larger d is, the more the node evolution converges to (1,1,1) the faster the rate of convergence. The positive service of the platform can bring good reputational benefits, and as F_h gets higher, the rate of evolution of individual nodes to 1 gradually increases faster. Likewise, negative services also bring reputational losses to the platform, and the higher the F_l, the faster each node converges to 1. With blockchain technology, reasonable and effective penalties and incentives can promote positive service of platform nodes and capacity sharing of enterprise nodes.

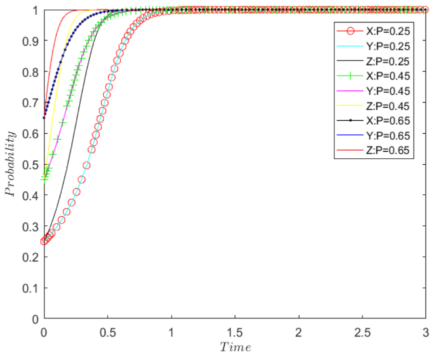


Fig. 9 Impact of P

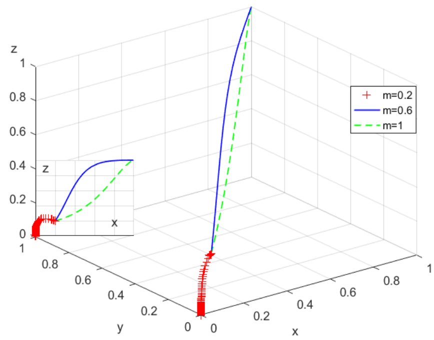


Fig. 10 Impact of m

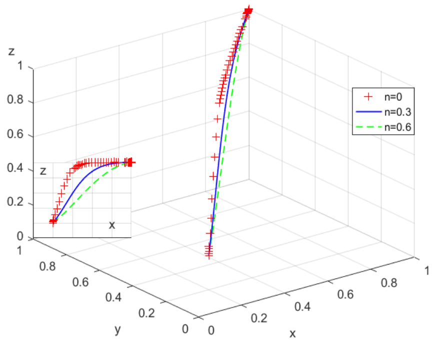


Fig. 11 Impact of n

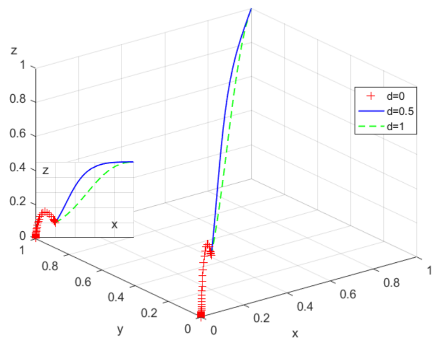


Fig. 12 Impact of d

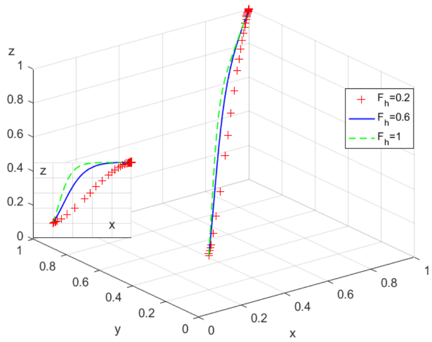


Fig. 13 Impact of F_h

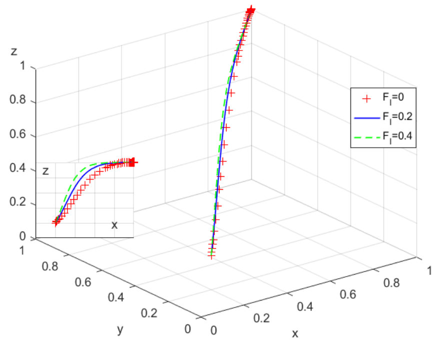


Fig. 14 Impact of F_l

6. Conclusion and management recommendations

6.1 Conclusion

From the results of the above analysis, it can be seen that:

- The game subjects will not change the evolutionary results under the initial value setting, no matter what the proportion of strategy selection is, if the relevant parameters are changed, it will have an impact on the strategy selection of the three subjects.
- Incentive and penalty coefficients are set under blockchain technology for rewarding the sharing party and punishing the non-sharing party. The incentive coefficient under blockchain technology has a positive impact on the sharing behaviour of manufacturing enterprises, and the penalty coefficient has a negative impact on the sharing behaviour of manufacturing enterprises.
- The digital benefit coefficient affects the behaviour of manufacturing enterprises, too high or too low is not conducive to the choice of manufacturing enterprises, and an appropriate digital benefit coefficient will promote the sharing behaviour of manufacturing enterprises.

- The platform under blockchain technology introduces a regulatory mechanism combined with smart contracts to constrain the behaviour of enterprises and the platform. Both positive regulation and negative regulation coefficients have a positive impact on the behaviour of the three parties.

6.2 Recommendation

Based on the above research findings, the following recommendations are made. On the one hand, the alliance nodes under blockchain technology can appropriately optimize the incentive mode and penalty mode, and find suitable allocation coefficients to maximize the incentive for the capacity-sharing behaviour of both supply and demand sides and promote the cooperation between them; at the same time, the third-party platform should actively supervise and strengthen its supervisory capacity to avoid the emergence of malicious pigeonholing behaviour, improve the effective cooperation between supply and demand sides and promote the long-term development of the manufacturing industry. In addition, both supply and demand sides should maintain good integrity to help create a good market environment, thus promoting the benign development of the manufacturing industry.

6.3 Shortcomings

The article uses an evolutionary game approach to investigate the capacity sharing behaviour of manufacturing firms and uses numerical simulations in MATLAB to analyse the effect of changes in different parameters on the evolutionary outcome. However, the article does not consider the effects of other aspects such as product quality, delivery time, human emotion, social development and environment on the game behaviour of capacity sharing, which can be further explored by introducing prospect theory in the future, and I am currently conducting related research.

Acknowledgement

Thanks to the National Social Science Foundation of China for funding (2022BJY021).

References

[1] Dai, H.-N., Zheng, Z., Zhang, Y. (2019). Blockchain for Internet of things: A survey, *IEEE Internet of Things Journal*, Vol. 6, No. 5, 8076-8094, doi: 10.1109/JIOT.2019.2920987.

[2] Hariri, R.H., Fredericks, E.M., Bowers, K.M. (2019). Uncertainty in big data analytics: Survey, opportunities, and challenges, *Journal of Big Data*, Vol. 6, No. 1, Article No. 44, doi: 10.1186/s40537-019-0206-3.

[3] Zhang, C., Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects, *Journal of Industrial Information Integration*, Vol. 23, Article No. 100224, doi: 10.1016/j.jii.2021.100224.

[4] Abhishek, V., Guajardo, J.A., Zhang, Z. (2021). Business models in the sharing economy: Manufacturing durable goods in the presence of peer-to-peer rental markets, *Information Systems Research*, Vol. 32, No. 4, 1450-1469, doi: 10.1287/isre.2021.1034.

[5] Ojstersek, R., Javernik, A., Buchmeister, B. (2021). The impact of the collaborative workplace on the production system capacity: Simulation modelling vs. real-world application approach, *Advances in Production Engineering & Management,* Vol. 16, No. 4, 431-442, doi: 10.14743/apem2021.4.411.

[6] Guo, L., Wu, X. (2018). Capacity sharing between competitors, *Management Science*, Vol. 64, No. 8, 3554-3573, doi: 10.1287/mnsc.2017.2796.

[7] Chen, J., Wang, X., Chu, Z. (2020). Capacity sharing, product differentiation and welfare, *Economic Research-Ekonomska Istraživanja*, Vol. 33, No. 1, 107-123, doi: 10.1080/1331677X.2019.1710234.

[8] Fang, D., Wang, J. (2020). Horizontal capacity sharing between asymmetric competitors, *Omega*, Vol. 97, Article No. 102109, doi: 10.1016/j.omega.2019.102109.

[9] Nofer, M., Gomber, P., Hinz, O., Schiereck, D. (2017). Blockchain, *Business & Information Systems Engineering*, Vol. 59, No. 3, 183-187, doi: 10.1007/s12599-017-0467-3.

[10] Berdik, D., Otoum, S., Schmidt, N., Porter, D., Jararweh, Y. (2021). A survey on blockchain for information systems management and security, *Information Processing & Management*, Vol. 58, No. 1, Article No. 102397, doi: 10.1016/j.ipm.2020.102397.

[11] Karamchandani, A., Srivastava, S.K., Kumar, S., Srivastava, A. (2021). Analysing perceived role of blockchain technology in SCM context for the manufacturing industry, *International Journal of Production Research*, Vol. 59, No. 11, 3398-3429, doi: 10.1080/00207543.2021.1883761.

[12] Catalini, C., Gans, J.S. (2020). Some simple economics of the blockchain, *Communications of the ACM*, Vol. 63, No. 7, 80-90, doi: 10.1145/3359552.

[13] Wang, X.L. (2021). Game-based hybrid particle swarm optimization of job-shop production control, *International Journal of Simulation Modelling*, Vol. 20, No. 2, 398-409, doi: 10.2507/IJSIMM20-2-CO9.

[14] Xiao, M.; Tian, Z.Y. (2022). Evolutionary game analysis of company collaborative strategy in cloud manufacturing platform environment, *Advances in Production Engineering & Management*, Vol. 17, No. 3, 295-310, doi: 10.14743/apem2022.3.437.

[15] Xiong, H., Dalhaus, T., Wang, P., Huang, J. (2020). Blockchain technology for agriculture: Applications and rationale, *Frontiers in Blockchain,* Vol. 3, No. 7, doi: 10.3389/fbloc.2020.00007.

[16] Yang, S.Y, Tan, C. (2022). Blockchain-based collaborative management of job shop supply chain, *International Journal of Simulation Modelling,* Vol. 21, No. 2, 364-374, doi: 10.2507/IJSIMM21-2-CO10.

[17] Chang, V., Baudier, P., Zhang, H., Xu, Q., Zhang, J., Arami, M. (2020). How blockchain can impact financial services– The overview, challenges and recommendations from expert interviewees, *Technological Forecasting and Social Change*, Vol. 158, Article No. 120166, doi: 10.1016/j.techfore.2020.120166.

[18] Yan, K., Cui, L., Zhang, H., Liu, S., Zuo, M. (2022). Supply chain information coordination based on blockchain technology: A comparative study with the traditional approach, *Advances in Production Engineering & Management*, Vol. 17, No. 1, 5-15, doi: 10.14743/apem2022.1.417.

[19] Abeyratne, S.A., Monfared, R.P. (2016). Blockchain ready manufacturing supply chain using distributed ledger, *International Journal of Research in Engineering and Technology*, Vol. 5, No. 9, 1-10, doi: 10.15623/ijret.2016.0509001.

[20] Leng, J., Ruan, G., Jiang, P., Xu, K., Liu, Q., Zhou, X., Liu, C. (2020). Blockchain-empowered sustainable manufacturing and product lifecycle management in industry 4.0: A survey, *Renewable and Sustainable Energy Reviews*, Vol. 132, Article No. 110112, doi: 10.1016/j.rser.2020.110112.

[21] Li, J., Maiti, A., Springer, M., Gray, T. (2020). Blockchain for supply chain quality management: Challenges and opportunities in context of open manufacturing and industrial internet of things, *International Journal of Computer Integrated Manufacturing*, Vol. 33, No. 12, 1321-1355, doi: 10.1080/0951192X.2020.1815853.

[22] Tian, C. (2020). Study of innovation manufacturing capacity sharing mode based on blockchain, *Science and Technology Management Research*, Vol. 40, No. 11, 9-14, doi:10.3969/j.issn.1000-7695.2020.11.002.

[23] Wang, T., Zhang, H. (2022). Evolutionary game study on manufacturing capacity sharing based on block chain, In: *Proceedings of 2022 3rd International Conference on Electronics, Communications and Information Technology (CECIT),* Sanya, China, 373-378, doi: 10.1109/CECIT58139.2022.00071.

[24] Hao, J.-Q., Zhao, D.-Z. (2021). Tripartite-players evolutionary game analysis of the manufacturing capacity sharing in the environment of sharing economy, *Operations Research and Management Science*, Vol. 30, No. 2, 1-7.

[25] Xin, Y.-L. (2020). Evolutionary game analysis of manufacturing enterprise capacity sharing strategy from the perspective of competition and cooperation, *Technology and Innovation Management,* No. 4, 375-379, doi: 10.14090/j.cnki.jscx.2020.0409.

[26] Zhu, L.-L., Rong, J.-M., Zhang, S.-Y. (2021). Three-party evolutionary game and simulation analysis of drug quality supervision under the government reward and punishment mechanism, *Chinese Journal of Management Science*, Vol. 2021, No. 11, 55-67, doi: 10.16381/j.cnki.issn1003-207x.2019.0481.

APEM journal

Advances in Production Engineering & Management ISSN 1854-6250
Volume 18 | Number 2 | June 2023 | pp 237–249 Journal home: apem-journal.org
https://doi.org/10.14743/apem2023.2.470 Original scientific paper

Enhancing automated defect detection through sequential clustering and classification: An industrial case study using the Sine-Cosine Algorithm, Possibilistic Fuzzy *c*-means, and Artificial Neural Network

Nguyen, T.P.Q.[a,*], Yang, C.L.[b], Le, M.D.[c], Nguyen, T.T.[d], Luu, M.T.[e]

[a]Faculty of Project Management, The University of Danang – University of Science and Technology, Danang, Vietnam
[b]Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan (ROC)
[c]Faculty of Transportation Mechanical Engineering, The University of Danang – University of Science and Technology, Danang, Vietnam
[d]Faculty of Computer Science and Engineering, University of Technology, Vietnam
[e]Nghia Nippers Corporation, Vietnam

ABSTRACT

Most existing inspection models solely classify defects as either good or bad, focusing primarily on separating flaws from perfect ones. The sequential clustering and classification technique (SCC) is used in this work to not only identify and categorize the defects but also investigate their root causes. Conventional clustering techniques like *k*-means, fuzzy *c*-means, and self-organizing map are employed in the first stage to find the defects in the finished products. Then, a novel clustering method, that combines a sine-cosine algorithm and possibilistic fuzzy *c*-means (SCA-PFCM), is proposed to classify the detected defects into multiple groups to identify the defect categories and analyze the root causes of failures. In the second stage, the ground truth labels taken from the clustering technique are used to construct an automated inspection system using back propagation neural networks (BPNN). The proposed approach is applicable for detecting and identifying the causes of errors in manufacturing industry. This study applies a case study in nipper manufacture. The SCA-PFCM algorithm can detect 97 % of defects and classify them into four types while BPNN shows a predicted accuracy of up to 96 %. Additionally, an automated inspection system is developed to reduce the time and cost of the inspection process.

ARTICLE INFO

Keywords:
Back Propagation Neural Network;
Clustering;
Classification;
Combined SCA-PFCM;
Defect detection;
Nipper manufacturing;
Possibilistic Fuzzy *c*-means (PFCM);
Root cause analysis;
Sine-Cosine Algorithm (SCA)

**Corresponding author:*
ntpquyen@dut.udn.vn
(Nguyen, T.P.Q)

Article history:
Received 24 April 2023
Revised 12 June 2023
Accepted 23 June 2023

1. Introduction

The rapid rise of both information and operational technology is accelerating the transition from traditional manufacturing to smart factories. A smart factory typically relies on modern information and communication technology, with all of the factory's components being smartly integrated and operated [1]. A smart factory needs to optimize production conditions with a minimum of resources and time. Defects are one of the primary causes of high production costs. Therefore, the inspection system for defect detection is crucial in smart manufacturing, which necessitates the quick detection of defects and categorizes defect types. Effective defect categorization aids in determining and analyzing the type of failure that occurs based on the process conditions [2]. Besides, the defect types can be used to diagnose equipment failure and analyze its root causes. Moreover, an effective inspection system can raise customer satisfaction levels by preventing them from obtaining substandard goods.

Many manufacturing factories still operate the visual and sampling inspection manually. Sampling inspections cannot catch all the defects like a thorough inspection can. Manual inspection may lead to inconsistent results, expensive costs, and time-consuming [3]. Enterprises expect quick defect detection and root cause investigation to provide timely regulation of production line failures [4]. A more effective and reliable automated defect inspection system is required to overcome the drawbacks of sampling and manual inspection in manufacturing.

Generally, a data analysis framework to detect the defects and investigate their causes is necessary. However, the majority of the research that has been done so far has solely focused on identifying defects or defect prediction models, which can result in dividing the finished products into two groups: 1) defects, and 2) non-defective ones [5]. In practice, the manager may not only need to detect failure or an abnormal state in the manufacturing process but also identify the causes of defects in the routine execution of the production plan to make timely adjustments and corrections [6]. This process is known as root cause analysis which aims to determine the underlying cause of an issue and the measures required to solve it.

A novel model of sequential clustering and classification-based genetic algorithm (NSGAII-SCC) has recently been proposed by Yang and Quyen to investigate the hidden structure of data and to identify the features correlated with the explored patterns [7]. The NSGAII-SCC framework is applied to analyze the point of sale (POS) data for a chain of bakeries in China [8]. The bakery store is partitioned into several clusters using a clustering technique and classification is then used to investigate the factors that contribute to the partition process of each store cluster. Kuo *et al.* presented an extension of the NSGAII-SCC by combining a deep learning and multi-objective sine-cosine algorithm (Deep MOSCA-SCC) [9]. The Deep MOSCA-SCC algorithm employed a deep clustering technique that combined auto-encoder and k-means to improve clustering performance. Generally, the SCC approach is useful for defect detection and classification, analyzing its root causes as well as establishing an automatic detect defection model in the above analysis.

Thus, this study focuses on developing a model-based SCC method to detect the defects and analyze their root causes. In the first stage, the proposed model-based SCC method first employs clustering methods to detect the defects in the finished product. Thereafter, the defects are classified into several groups with similar properties to identify the defect categories. Most of the research used traditional clustering techniques like k-means algorithm[10] or fuzzy c-means (FCM) algorithm [11] for industrial applications since they are simple to implement and interpret the result. However, the clustering result was also comparable. In the first stage, several conventional clustering techniques, i.e., k-means, FCM, self-organizing map (SOM) [12], and DBSCAN [13], are employed to detect the defects from the flawless ones since this process is not very complicated due to the predetermined number of clusters k is known as two labels (defect or non-defect). To determine the defect categories and analyze the root causes of failures, a novel clustering method that combines a sine-cosine algorithm and possibilistic fuzzy c-means (SCA-PFCM), is proposed. In the second stage, a back propagation neural network (BPNN) [14] is employed to develop an automatic defect detection model based on the ground truth labels adopted from the clustering technique. In contrast to existing SCC methods, the proposed inspection model-based SCC contributes by combining innovative clustering and classification techniques, namely SCA-PFCM and BPNN, whereas the original NSGAII-SCC suggested an integration of traditional k-means and decision tree classifier. The research result is then applied to an inspection system of a nipper manufacturing factory in Vietnam.

The paper is arranged as follows. The review of sequential clustering and classification method is shown in Section 2. Section 3 describes the production and inspection process in nail nipper manufacturing. The methodology is presented in Section 4. Section 5 illustrates the result analysis of a case study. The concluding remarks and future research direction come in Section 6.

2. Methods and materials

2.1 Review of sequential clustering and classification method

Since clustering is frequently carried out without knowledge regarding the membership of data instances to predetermined labels, it is commonly referred to as a technique for unsupervised learning [15]. The goal of clustering is exploratory the structure of the dataset. The clustering process divides objects in a given dataset into distinct groups or clusters, with each cluster consisting of objects that are similar to each other in that cluster but dissimilar to objects in other clusters. On the contrary, classification is a supervised approach that assigns categories or labels to data records based on previously collected information [16]. The classification process consists of two steps. The first step is to develop a training model based on the classification rules using the given data, which contains a set of attributes and their corresponding outcomes. The training model is then used in the second step to forecast the labels for incoming unknown data. The complexity of the large number of features utilized as a classifier is what makes these processes challenging [17].

Yang and Nguyen [7] proposed the SCC framework, which combines clustering and classification sequentially on two different datasets. The SCC framework can investigate the hidden structure of data and discover the features that are relevant to the explored patterns. Fig. 1 illustrates the SCC framework. From the original dataset, the target interests were identified and these features were then used for discovering data patterns by the clustering algorithm. The remaining dataset contained the relevant features that are correlated with the target features. A clustering algorithm was implemented on the dataset containing target features to explore the data patterns. The result of clustering algorithm was embedded into the training process of the classification model. Two popular clustering algorithms, k-means, and hierarchical clustering, were used to perform clustering. The classification algorithm employed decision trees [18], artificial neural networks (ANN) [19], k-nearest neighbor (KNN) [20], and support vector machine (SVM) [21]. As a result, there are eight clustering-classification algorithm combinations built from two clustering methods and four classification methods. Besides, as combining heuristic techniques and machine learning become more prevalent [22, 23], the SCC framework combined with NSGAII to deal with two objective functions from clustering and classification tasks. Because the number of clusters (k) is not predetermined, the SCC framework implemented the various values of k (from 2 to 10) and finally selected the solution based on the Pareto front. For each value of k, multiple solutions are selected from the first Pareto front to balance two objective functions of clustering and classification tasks. Performance was superior to the others when k-means clustering and decision tree classification were combined.

The Deep MOSCA-SCC algorithm [9] is also an SCC approach that employed deep clustering to enhance the clustering compactness and classification accuracy. Herein, MOSCA [24], which is a simple and easy-to-implement method, was applied to exploit the optimal result for the SCC approach. The utilization of MOSCA in the Deep MOSCA-SCC algorithm was expected to speed up computation compared to NSGAII. However, the experimental results demonstrated that there was no benefit in Deep MOSCA-SCC computational time due to the use of autoencoder, which increased computational time.

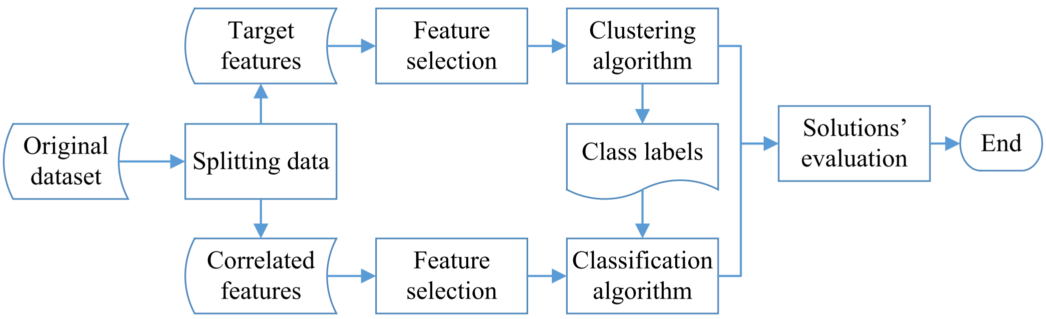


Fig. 1 The SCC approach

2.2 Proposed methodology

The manual examination procedure necessitates a significant amount of human labor and expenditure. Industry 4.0 is having an impact on nail manufacturing, with automated defect detection reducing inspection time and human labor. Automatic visual inspection, which plays a key part in the quality inspection process of smart manufacturing, is a promising technology in this scenario. The SCC technique is used for algorithm development of an automatic visual inspection in nipper manufacturing in this study.

This study utilizes the SCC method [7] to not only detect the defects from the flawless ones but also classify the defects into several groups with similar attributes for root cause analysis in the first stage. An automated defect detection system is developed in the second stage. The result of the first stage is used to train the classification model for automatic defect detection. The methodology framework is shown in Fig. 2.

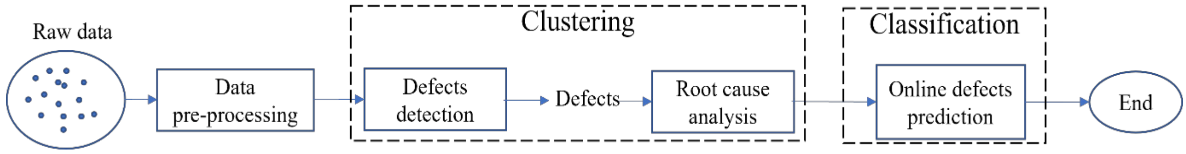


Fig. 2 Methodology Framework

Stage 1: Defect detection and root cause analysis using clustering technique

Detecting defects from the finished products

In this stage, the clustering techniques are employed for defect detection and root cause analysis in nipper manufacturing. The dataset is collected from a nipper factory in Vietnam. The Department of quality control provides historical data on finished items, which is only divided into two categories: defective products and non-defective products. Thus, this study first uses the clustering technique to detect the defects from the flawless ones. The number of clusters is defined as $k = 2$ and the class labels of all data instances are predetermined based on the historical data. k-mean, FCM, SOM, and DBSCAN are employed to classify the defects and non-defected items in the first stage since they are in the group of popular clustering methods applied for industry applications [25, 26]. Clustering accuracy ($Clus\_Acc$) is used to evaluate the result of this step. The $Clus\_Acc$ is calculated as follows [27]:

$$Clus\_Acc = \frac{\sum_{i=1}^{k} b_i}{n}, \tag{1}$$

where b_i is the number of instances that are classified in a corrected cluster, n is the number of total data instances.

Proposed SCA-PFCM for identifying the defect categories

The defects revealed in the first step are classified into several types of defects that have similar properties to analyze the root causes of defects. Herein, k is unknown. Thus, this study proposes a novel clustering method, i.e., SCA-PFCM, which not only automatically determines the optimal k but also simultaneously partitions the clusters to reveal data patterns corresponding to the selected k. The proposed SCA-PFCM inherited the approach of identifying the optimal number of clusters in an automatic fuzzy clustering method proposed by Nguyen and Kuo [28], which deals with automatic clustering for categorical data. However, the dataset from the nipper industry contains numerical features. Thus, the distance measure in the clustering procedure is changed to be appropriate for the data features. Besides, SCA is employed since it is a simple and straightforward algorithm with competitive performance as compared with GA, and PSO [29]. The sine and cosine functions are utilized to update the particle's position in the SCA algorithm as follows:

$$X_i^{(t+1)} = X_i^t + r_1 \times \sin(r_2) \times \left|r_3 P_i^t - X_i^t\right|,\ r_4 < 0.5 \tag{2}$$

$$X_i^{(t+1)} = X_i^t + r_1 \times \cos(r_2) \times \left|r_3 P_i^t - X_i^t\right|,\ r_4 \geq 0.5 \tag{3}$$

where X_i^t *and* $X_i^{(t+1)}$ is the position of individual i at iteration t and $t+1$, respectively. P_i^t is the optimal position at the current iteration t. r_1 is a control number that is calculated as:

$$r_1 = \mu - t\frac{\mu}{T} \tag{4}$$

while r_2 is a number randomly chosen between 0 and 2π, whereas $r_3, r_4 \in (0,2)$. μ is a constant.

The SCA-PFCM algorithm's process is explained as follows:

Step 1: Identify the maximum k (k_{max})
A local density-based approach named RECOME [30] is employed to determine k_{max}. The cluster center typically has higher local density and is surrounded by neighbors with lower local densities. If the number of high-density centers is discovered, the k_{max} can be obtained. The procedure to determine k_{max} is shown in Fig. 3.

Step 2: Set up parameters for SCA and PFCM algorithms.

Step 3: Initialization: Randomly generate the initial population. A solution representation consists of two parts. The first part is used to define the k. A control element C_p is used to determine k in each particle where $k_{min} \leq C_p \leq k_{max}$. k_{min} is set at 2. To determine C_p, we randomly generate a vector $C = C_1, C_2, \dots, C_j, \dots, C_{k_{max}}$ in the range of [0, 1]. C_p is determined by counting the elements in C that are greater than 0.5: C_p = count (C|C_j ≥ 0.5). The second part of a particle is the cluster center corresponding to the value of k in the first part.

Step 4: Calculate fitness: The PFCM algorithm is implemented for each particle in the population. The fitness value is obtained through the objective function of PFMC algorithm.

Step 5: Identify the optimal solution P_i^t at iteration *t.*

Step 6: Compute r_1 using Eq. 4 and randomly generate r_2 r_3, and r_4.

Step 7: Update the cluster centers using Eq. 2 and Eq. 3.

Step 8: Return to step 4 until the stopping condition is met.

Input: a given dataset X, $N(x)$ is the K nearest neighbors set of x in X.

Output: k_{max}

- For each instance $x \in X$, and $N(x)$ is the K nearest neighbors set of x in X:
    - Calculate K nearest Neighbor Kernel Density (NKD):

$$\rho(x) = \theta \sum_{y \in N(x)} \exp\left(-\frac{d(x,y)}{\sigma}\right)$$

    - Caculate relative K nearest Neighbor Kernel Density (RNKD):

$$\rho^*(x) = \frac{\rho(x)}{\max_{y \in N(x) \cup \{x\}} \rho(y)}$$

- Define core instances: $O = \{x | x \in X, \rho^*(x) = 1\}$.
- Determine Higher Density Nearest-neighbor (HDN)-$\pi(x)$:

$$\pi(x) = argmin\,\{d(x,y)\}, where\ \rho(y) > \rho(x), y \in X.$$

- Construct a directed graph $G = (X, A)$, where $A = \{\langle x, \pi(x)\rangle | x \in X \backslash O\}$.
- Find atom cluster: for a core instance C_o in O
    - The atom cluster $C_a = \{C_o\} \cup \{x \in XO | x\ is\ connected\ to\ C_o\ in\ G\}$.
- k_{max} is defined by number of atom clusters

Fig. 3 Procedure to determine k_{max} [28]

The result in this stage, which includes the class labels and the number of clusters for defects, is used to train the model in the second stage.

Stage 2: Automatic defect detection using back propagation neural network (BPNN)

BPNN [14] is a popular branch of ANN that contains a multi-layer feedforward network trained by the error BP technique. A BPNN consists of three different layer types: input, hidden, and output layers. The hidden layer acts as a link for signal forward propagation connecting the input and output layers. If the output layer is unable to produce the required results, the error BP will transmit an error signal from the output nodes back via the hidden layer to the input nodes before repeating the signal forward propagation.

This study employs BPNN to develop an automated defect prediction model. The labels that are revealed by clustering techniques in the first stage are used to train the BPNN model. The number of neurons in the input layer is identified by the number of data features that significantly impact the output. The output layer nodes depend on the defect categories that are partitioned in the first stage. It is calculated by multiplying the number of defect types by one to represent the defect categories and good products. The performance of the BP network is significantly influenced by the number of hidden layers and hidden layer nodes. A single hidden layer or two hidden layers can be considered in the design of BPNN model depending on its predictive performance [31]. Besides, better performance can be achieved with more hidden layer nodes, although a lengthy training period may result. Typically, an empirical formula is used to estimate the number of hidden layer nodes. The empirical formula is displayed as follows [32]:

$$H = \sqrt{I + O} + a, \tag{5}$$

where I, H, and O are the number of nodes in the input, hidden, and output layers, respectively, a is a constant number in the range [0, 10]. Cross-entropy is selected as the loss function, which is calculated as follows [33]:

$$E = -\sum_{i=1}^{N}(y_i * log(\hat{y}_i) + (1 - y_i) * log(1 - \hat{y}_i)), \tag{6}$$

where y_i is the target value, $\hat{y}_i$ is predicted value. The structure of BPNN that contains two hidden layers for automated defect prediction is shown in Fig. 4.

The proposed method is applicable for detecting and identifying the causes of errors in industrial manufacturing. There are two ways that the input data can be gathered from the industry: sensor signal data or image data. The proposed method in its current form can work well for signal data. If the input is the image data, it can be preprocessed to extract the features from the image before the proposed method is implemented. This paper analyzes a case study in nipper manufacturing. An automatic defect detection system is developed in Section 3.3 in which the product images are taken on the production line. The user can adjust the model to match the type of input data.

The nipper manufacturing industry in Vietnam, which is selected as a case study for this research, will be covered in the next subsection, along with the production and inspection processes.

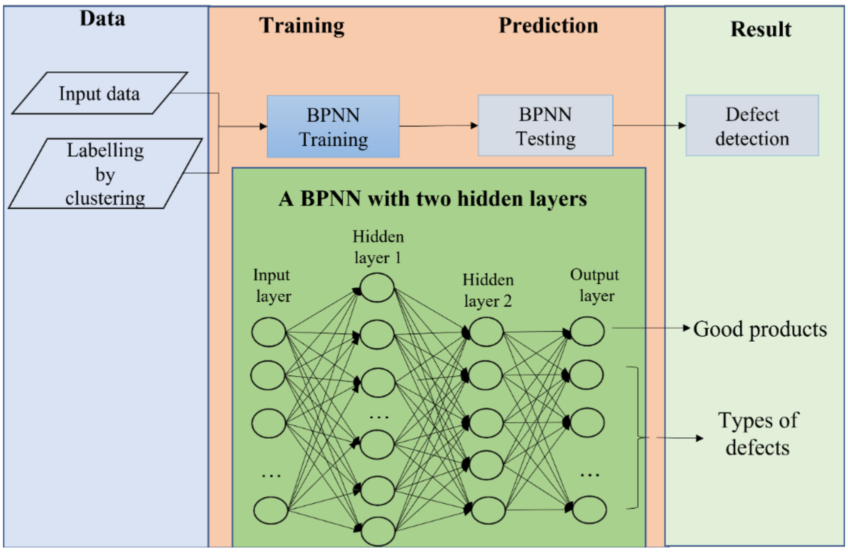


Fig. 4 The structure of BPNN model for automated defect prediction

2.3 Production and inspection processes in nipper manufacturing

This section introduces the production and inspection processes in nipper manufacturing. There are two types of nippers: cuticle nippers and nail nippers. The cuticle nipper usually has a smaller jaw and handle, as compared to the nail nipper. Besides, the cuticle nipper also has a lighter weight than the nail nipper. A nipper consists of five main components: jaw, joint, spring, spring pin, and handles. The jaw is designed with exact blade alignment and precision sharpening. For the cuticle nipper, there is different jaw size such as J12, J14, and J16 which are corresponding to the blade sizes are 4.5-5.5 mm, 5.5-6.5 mm, and 6.5-7.5 mm, respectively. In contrast, the nail nipper only has one jaw size of 6.5-8.0 mm. The joints aim to transmit the cutting force from the handles to the two jaws when the user squeezes them. Box joint is designed to warranty it is durable. Springs are used to assist (push and elastic) the user when squeezing the two handles. The pin is used to hold the spring with the handle. Besides, the company name or logo is also etched on the handle.

The manufacturing process of a nipper is described in Fig. 5. There are four main stages in the whole production process. The first stage is to make a mold from the mold blueprint. In the pressing stage, several processes, such as cutting, stamping for shaping, bending the blade edges, annealing mill, and grinding the surface, are continuously carried out. The output of this stage is a shaped nipper. The process is continued in the pre-processing stage which includes the stamping and drilling assembly, thermal process, shaping of the jaw, polishing of the handle, and plating. Thereafter, the preliminary nipper is transferred to the finishing stage with the following processes: installing springs, printing laser, polishing the jaw, sharpening the blade edge, and cleaning the finished product. There is an inspection process to check product quality at each stage. Then, the inspection process is applied to the finished product. The criteria for inspection consist of sharpness, the shape of the jaw, logo printing, length of the blade, and handle length. The inspection process is performed manually and is costly. Moreover, the number of products per day is also huge. Thus, sampling inspection is currently used with a sample of 10 % to 20 % of finished products randomly taken for inspection. The sampling inspection process cannot guarantee the quality of all the finished products.

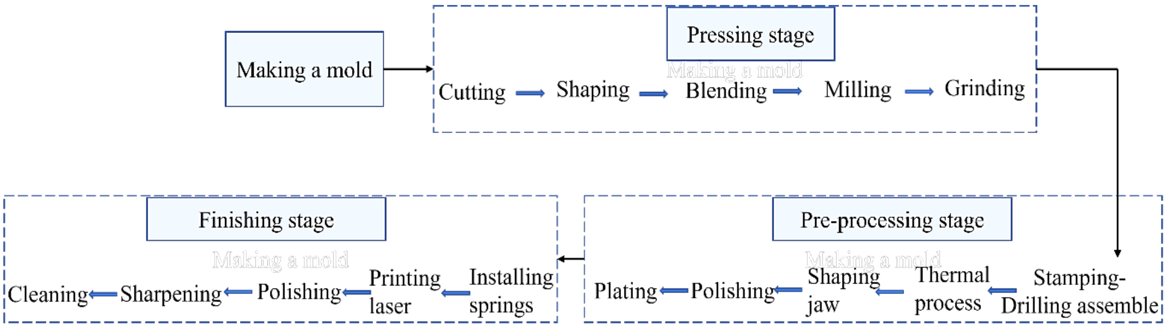


Fig. 5 The manufacturing process of a nipper

3. Case study: Result and discussion

3.1 Data collection and parameter setting

The experiment uses the historical data gathered from a nipper manufacturing factory in Vietnam. According to the company's current production needs, a total of 15000 nippers of various types must be produced per day. As a result, the amount of data recorded day by day becomes very huge to process. This study analyzes three different types of nail nippers, coded D555, D401, and D363, respectively. D555 includes 84320 data instances, whereas D401 and D363 have 48384 and 33690 data instances. Each data instance has the 12 features such as nipper length, handle, blade length, width and thickness of nipper handles, thickness of two nipper bearings, and so on.

In the first stage, k-means, FCM, SOM, and DBSCAN are used to classify the defects from the flawless ones. The number of clusters is predetermined as $k = 2$. Clustering accuracy is employed to evaluate the result. Then the discovered defects are used to perform clustering to determine the defect types by using the proposed SCA-PFCM algorithm. To compare the optimal k obtained from the SCA-PFCM, the elbow method-based k-means is employed in which the sum of squared error (SSE) is considered as a performance indicator. SSE is calculated as follows:

$$SSE = \sum_{j=1}^{k} \sum_{x \in C_j} dist(x, m_j)^2 \tag{7}$$

The parameter setting for BPNN model will be discussed in more detail in the next subsection after getting the result of the first stage.

3.2 Result analysis

First-stage experimental findings

Clustering methods are first used to detect the defects in the finished product. The k-means, FCM, SOM, and DBSCAN algorithms are implemented on the three datasets. Each algorithm is executed 30 times for each tested dataset. Table 1 shows the clustering accuracy based on the average values of 30 runs and the standard deviation. It is not much different from the clustering accuracy by k-means and FCM. The DBSCAN algorithm is slightly better since its accuracy dominates the other algorithms for all tested data. Thus, the clustering result of DBSCAN is used in the next step. Table 2 shows the number of defective and non-defective products detected by DBSCAN method.

Table 1 Clustering accuracy of the process to classify defects from flawless ones

Dataset	k-means	FCM	SOM	DBSCAN
D555	0.912 ± 0.001	0.904 ± 0.003	0.896 ± 0.003	0.913 ± 0.001
D401	0.938 ± 0.001	0.940 ± 0.002	0.913 ± 0.002	0.942 ± 0.002
D363	0.953 ± 0.001	0.951 ± 0.001	0.948 ± 0.002	0.965 ± 0.001

Table 2 Clustering results in terms of class labels

Dataset	Number of instances		
	Total	Non-defective	Defective
D555	84320	46873	37447
D401	48484	28245	20239
D363	33690	18970	14720

To analyze the root cause of failures, the defects are classified based on their features. The number of clusters, which is represented by how many types of defects, is unknown. Elbow method-based k-means clustering is applied. Fig. 6 shows the SSE plot for various k in the D555, D401, and D363 datasets, respectively. The result is quite consistent since the optimal k is selected at $k = 4$, which means that there are 4 types of defects.

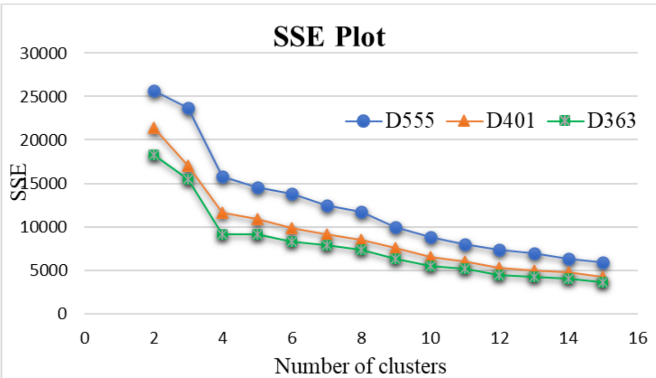


Fig. 6 Identify the number of defect types based on the elbow method

Besides, the number of defect types is also determined using the proposed SCA-PFCM. The result is quite similar to the elbow-based k-means since four types of defects are determined. However, the performance in terms of SSE of the SCA-PFCM is smaller than that of the elbow-based k-means, as shown in Table 3. It means that the clustering result provided by SCA-PFCM is more compact. The number of defect types and the label of each defect instance obtained from the

SCA-PFCM are used in the second stage. The defect types can be listed as follows. The first type of defect is untight jaws since there is a clearance between two blades of the nipper that exceeds the gap standard. In the blade sharpening process, the craftsman leaves a gap that is allowed according to the standard. If they sharpen it a lot, there is a gap between the two blades of the nipper, and the length of the blade may be shorter. Most of the defects derive from this cause. The second and third types of defect are large gaps in the upper and lower gills of the nipper caused by the assembly process of the box joint. The four defect type belongs to the length of the nipper blade as it can be shorter or longer than its standard length due to the blade grinding process.

Table 3 Comparison of SCA-PFCM and elbow-based k-means in terms of SSE

Dataset	Optimal k	SSE	
		SCA-PFCM	Elbow method
D555	4	14233.89	15784.78
D401	4	10602.53	11649.34
D363	4	8321.94	9157.63

Experimental results in the second stage

The number of defect types and the label for each defect are employed to train the model in the second stage. This stage aims to not only detect the defects from the finished products but also classify the defects into certain groups that are identified in the first stage. Hence, there are five labels representing four types of defects and good product and their corresponding number of instances are presented in Table 4.

The parameter for BPNN model with a single hidden layer (denoted as BPNN-1) is set up as follows. First, the input layer nodes equal to 12 which are the significant non-linear relationship variables to represent product features. The output layer nodes are 5 including the defect types and good product. Based on Eq. 5, the number of neurons in the hidden layer is from 4 to 14. After several trial experiments, the hidden layer nodes are selected as 10, initial weight $w = 0.2$, learning rate $r = 0.1$, and inertia weight $c = 0.6$. Logsig and purelin are the activation functions for the hidden and output layers, respectively [31]. Regarding the BPNN model with two hidden layers (BPNN-2), the input and output layers are set up similarly to the BPNN-1 model. The number of neurons in the first hidden layer is also from 4 to 14 as determined by Eq. 5. This study eliminates half of the constant number in Eq. 5, which now ranges from 0 to 5, to shorten the computation time. Thus, there are 4 to 9 neurons in the second hidden layer. The activation functions and other parameters are similar to the BPNN-1 model. After a series of experimental trials, the first and second hidden layers are chosen to have 12 and 6 neurons, respectively, based on their predictive performance. Table 5 shows the selection of key parameters for BPNN models.

Table 4 Data properties for classification

Label	Types of defect	Number of instances		
		D555	D401	D363
1	Non-defective	46873	28245	18970
2	Untight jaws	19326	11017	6793
3	Large gap in the upper gill of nipper	8768	3790	3278
4	Large gap in the lower gill of nipper	6293	4032	3480
5	Length of nipper blade	3060	1400	1169

Table 5 Setting for BPNN model

Model	Number of neurons			Activation function	
	Input layer	Hidden layer	Output layer	Hidden layer	Output layer
BPNN-1	12	10	5	logsig	purelin
BPNN-2	12	12-6	5	logsig	purelin

To evaluate how well the BPNN model predicts, the confusion matrix is utilized [34]. Four predictive performance metrics are obtained from the confusion metrics, as follows:

$$Accuracy = (\text{TP+TN})/(\text{TP+FP+TN+FN}) \tag{8}$$

$$Recall = \text{TP}/(\text{TP+FN}) \tag{9}$$

$$Precision = TP/(TP + FP) \tag{10}$$

$$F\_measure = 2 \cdot Recall \cdot Precision/(Recall + Precision) \tag{11}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Table 6 shows the experimental result in terms of the confusion matrix of dataset D555 implemented by the BPNN-1 model. The overall accuracy is 94.6 % which is calculated by taking the average of the correctly predicted values in each class. The overall error rate correspondingly is 5.4 %. Besides, Kappa coefficient [35] is also used to evaluate the agreement between predicted values and truth class values. The kappa coefficient is computed as 93.2 % to represent a perfect agreement. Similarly, the confusion matrix of dataset D555 implemented by the BPNN-2 model is displayed in Table 7. The overall accuracy, overall error rate, and Kappa coefficient achieved by BPNN-2 are 96 %, 4 %, and 94.8 %, respectively, which are relatively higher than those achieved by BPNN-1 model. Table 8 summarizes the result comparison of the BPNN-1 and BPNN-2 models in the three tested datasets. The BPNN-2 model is significantly better than the BPNN-1 model in terms of accuracy and Kappa coefficient. Besides, the computational time on the product group of each tested dataset is also presented in Table 8. The BPNN-2 takes more time for the training model. Therefore, compared to the BPNN-1, its computing time is significantly longer. However, the result in Table 8 is shown for the product group that includes numerous objects. The highest computational times, while taking into account the average computational time for each product, are 0.052 and 0.045 for BPNN-2 and BPNN-1, respectively. The average computation time per product provided by the two models does not differ considerably while the accuracy of BPNN-2 is significantly better. Thus, the BPNN-2 model is used to develop an automatic defect inspection system presented in the next subsection.

Table 6 Confusion matrix of dataset D555 implemented by the BPNN-1 model

BPNN-1 Dataset D555		Actual				
		Non-defective (%)	Untight jaws (%)	Large gap in the upper gill (%)	Large gap in the lower gill (%)	Over-length of nipper blade (%)
Predicted	Non-defective	100	0	0	0	0
	Untight jaws	0	100	2	13	0
	Large gap in the upper gill	0	0	88	0	0
	Large gap in the lower gill	0	0	8	87	2
	Over-length of nipper blade	0	0	2	0	98

Table 7 Confusion matrix of dataset D555 implemented by the BPNN-2 model

BPNN-2 Dataset D555		Actual				
		Non-defective (%)	Untight jaws (%)	Large gap in the upper gill (%)	Large gap in the lower gill (%)	Over-length of nipper blade (%)
Predicted	Non-defective	100	0	0	0	0
	Untight jaws	0	100	0	10	0
	Large gap in the upper gill	0	0	93	0	0
	Large gap in the lower gill	0	0	5	90	3
	Over-length of nipper blade	0	0	2	0	97

Table 8 Result comparison

Dataset	D555		D401		D363	
	BPNN-1	BPNN-2	BPNN-1	BPNN-2	BPNN-1	BPNN-2
Overall accuracy (%)	94.6	96	93.5	95.2	92.8	96.3
Overall error rate (%)	5.4	4	6.5	4.8	7.2	3.7
Kappa Coefficient (%)	93.2	94.8	93.1	94.7	92.1	95.9
Computational time (s)	2440.5	2916.2	1795.5	2091.9	1529.2	1751.6

3.3 Automated defect inspection system

The BPNN-2 model whose results were validated in Section 5.2.2 is used in the automatic inspection system for nipper manufacturing. To transfer from manual inspection to the automatic inspection process, an advanced sensing system with smart camera sensors is installed. The system includes two industry cameras that are installed perpendicular to the conveyor. One camera is used to check the overall sizes of a nipper. Another one is put closer for enlarging to check the criteria related to the nipper's jaw. The cameras take pictures of the product from different angles. Then the system will analyze the images to extract the related feature to evaluate the product quality. There are a total of 12 features extracted from the image captured by from sensor camera, such as the nipper length, handle, blade length, the clearance between two nipper blades, width, and thickness of nipper handles, the thickness of two nipper bearings, and so on. Besides, the optical illumination system is installed under a transparent conveyor and shines in the opposite direction through the conveyor towards the camera.

The automatic visual inspection system in nipper manufacturing is displayed in Fig. 7. The system is operated as follows. Operator 1 puts the nippers on the conveyor in the correct direction. The system will adjust to make the nipper on the conveyor alignment. The sensor camera takes pictures automatically when the nipper passes through the inspection position. The image is transmitted to the processing system and the product features are extracted.

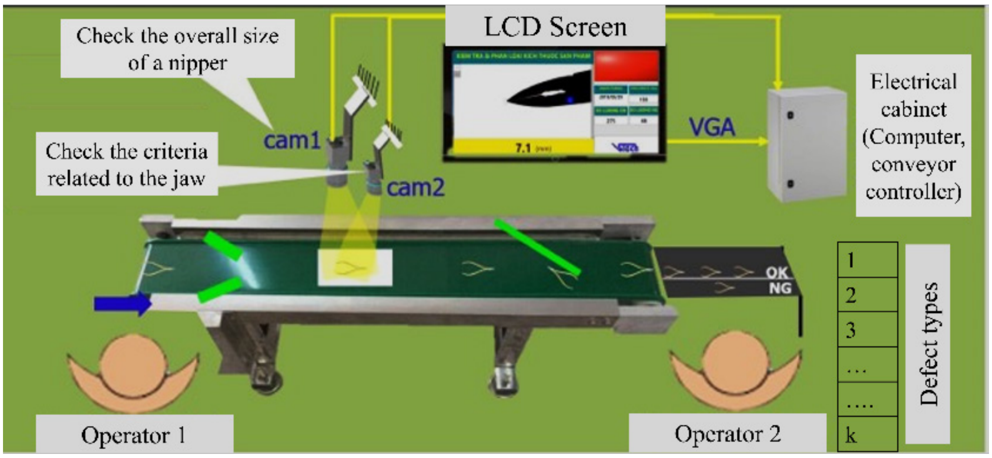


Fig. 7 The automatic inspection system in nipper manufacturing

The previous system at the company only detects a nipper to a defect or non-defective product. If a tested nipper is determined as a non-defective one, the product sorting mechanism is not active. Thus, the nipper will automatically transfer to the finished product tray. Otherwise, the product sorting mechanism is active to push a nipper on the other side of the conveyor and transfer it to the defect tray. However, this study considers not only detecting the defects but also finding the root cause of defects. Thus, the defects are classified into four groups based on their features as the result of the experiment. A classification model based on the clustering result in the first stage is developed to classify the tested nipper and determine exactly what type of defect the tested nipper is. The result is not only displayed on the screen but also has a sound warning to remind Operation 2 of the types of defects. Operator 2 has a responsibility to take the classified defect into the correct tray of defect types.

4. Conclusion

Defect detection is critical for quality control in the company. By detecting the defects in the finished products, the company can prevent sending out defects to customers. Classifying the defects into different categories based on their features also helps to determine the defect's causes. The management can utilize this information to make better decisions about addressing the defects or eliminating the factors that caused defects. Thus, the SCC method is utilized in this study to perform these works for the inspection process in the nipper factory in Vietnam. The DBSCAN clustering method is used to detect the defects in the finished products. Besides, a novel SCA-PFCM algorithm is proposed for defect categories. Then BPNN is employed to make an automatic inspection system. The result showed that the clustering and classification accuracy are all high to prove that the model is robust. The defects in the nipper were detected and classified into four different types based on the features of the defective products. Based on this result, an automatic defect detection model-based visual inspection was developed to help the factory improve the inspection process in terms of quality, time, and cost.

There are some limitations in the proposed method that can be enhance for further research. First, the clustering stage is critical important and affect the result of the classification stage since its result is embedded in the classification process. Thus, improving the clustering result is necessary. The proposed SCA-PFCM only considered within cluster distance as the objective function in the clustering process. Future research can investigate mul-objective function

This research can be extended to detecting the failures in the whole manufacturing process, not only in the finishing stage, as the recommendation by the manager. Besides, developing an algorithm to improve the quality of combining clustering and classification is also necessary.

Acknowledgment

This work was funded by Vingroup Joint Stock Company (Vingroup JSC) and supported by Vingroup Innovation Foundation (VINIF) under project code VINIF.2020.DA19. The support is much appreciated. The authors also thank Nghia Nippers Corp. for providing a dataset and support during the project.

References

[1] Yun, J.P., Shin, W.C., Koo, G., Kim, M.S., Lee, C., Lee, S.J. (2020). Automated defect inspection system for metal surfaces based on deep learning and data augmentation, *Journal of Manufacturing Systems*, Vol. 55, 317-324, doi: 10.1016/j.jmsy.2020.03.009.

[2] Wang, T., Chen, Y., Qiao, M., Snoussi, H. (2018). A fast and robust convolutional neural network-based defect detection model in product quality control, *International Journal of Advanced Manufacturing Technology*, Vol. 94, No. 9, 3465-3471, doi: 10.1007/s00170-017-0882-0.

[3] Ko, T., Lee, J.H., Cho, H., Cho, S., Lee, W., Lee, M. (2017). Machine learning-based anomaly detection via integration of manufacturing, inspection and after-sales service data, *Industrial Management & Data Systems*, Vol. 117, No. 5, 927-945, doi: 10.1108/IMDS-06-2016-0195.

[4] Yang, Z.J., Du, X.J., Chen, F., Chen, C.H., Tian, H.L., He, J.L. (2018). Change-point estimation for repairable systems combining bootstrap control charts and clustering analysis: Performance analysis and a case study, *Advances in Production Engineering & Management*, Vol. 13, No. 3, 307-320, doi: 10.14743/apem2018.3.292.

[5] Božič, J., Tabernik, D., Skočaj, D. (2021). Mixed supervision for surface-defect detection: From weakly to fully supervised learning, *Computers in Industry*, Vol. 129, Article No. 103459, doi: 10.1016/j.compind.2021.103459.

[6] Chongwatpol, J. (2015). Prognostic analysis of defects in manufacturing, *Industrial Management & Data Systems*, Vol. 115, No. 1, 64-87, doi: 10.1108/IMDS-05-2014-0158.

[7] Yang, C.-L., Quyen, N.T.P. (2018). Data analysis framework of sequential clustering and classification using non-dominated sorting genetic algorithm, *Applied Soft Computing*, Vol. 69, 704-718, doi: 10.1016/j.asoc.2017.12.019.

[8] Yang, C.-L., Nguyen, T.P.Q. (2022). Sequential clustering and classification approach to analyze sales performance of retail stores based on point-of-sale data, *International Journal of Information Technology & Decision Making,* Vol. 21, No. 3, 885-910, doi: 10.1142/S0219622022500079.

[9] Kuo, R.J., Setiawan, M.R., Nguyen, T.P.Q. (2022). Sequential clustering and classification using deep learning technique and multi-objective sine-cosine algorithm, *Computers & Industrial Engineering*, Vol. 173, Article No. 108695, doi: 10.1016/j.cie.2022.108695.

[10] Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 24, No. 7, 881-892, doi: 10.1109/TPAMI.2002.1017616.

[11] Bezdek, J.C., Ehrlich, R., Full, W. (1984). FCM: The fuzzy *c*-means clustering algorithm, *Computers & Geosciences*, Vol. 10, No. 2-3, 191-203, doi: 10.1016/0098-3004(84)90020-7.

[12] Vesanto, J., Alhoniemi, E. (2000). Clustering of the self-organizing map, *IEEE Transactions on Neural Networks,* Vol. 11, No. 3, 586-600, doi: 10.1109/72.846731.

[13] Ienco, D., Bordogna, G. (2018). Fuzzy extensions of the DBScan clustering algorithm, *Soft Computing*, Vol. 22, No. 5, 1719-1730, doi: 10.1007/s00500-016-2435-0.

[14] Goh, A.T.C. (1995). Back-propagation neural networks for modeling complex systems, *Artificial Intelligence in Engineering*, Vol. 9, No. 3, 143-151, doi: 10.1016/0954-1810(94)00011-S.
[15] Nguyen, T.P.Q., Kuo, R.J., Le, M.D., Nguyen, T.C., Le, T.H.A. (2022). Local search genetic algorithm-based possibilistic weighted fuzzy *c*-means for clustering mixed numerical and categorical data, *Neural Computing and Applications*, Vol. 34, No. 20, 18059-18074, doi: 10.1007/s00521-022-07411-1.
[16] Tan, P.N., Steinbach, M., Kumar, V. (2016). *Introduction to data mining,* Pearson Education, New Delhi, India.
[17] Zhang, Z.L., Wang, Y.F., Li, Y. (2019). Inventory control model based on multi-attribute material classification: An integrated grey-rough set and probabilistic neural network approach, *Advances in Production Engineering & Management*, Vol. 14, No. 1, 93-111, doi: 10.14743/apem2019.1.314.
[18] Zhang, Y.D., Liao, L., Yu, Q., Ma, W.G., Li, K.H. (2021). Using the gradient boosting decision tree (GBDT) algorithm for a train delay prediction model considering the delay propagation feature, *Advances in Production Engineering & Management*, Vol. 16, No. 3, 285-296, doi: 10.14743/apem2021.3.400.
[19] Tao, X., Zhang, D., Ma, W., Liu, X., Xu, D. (2018). Automatic metallic surface defect detection and recognition with convolutional neural networks, *Applied Sciences,* Vol. 8, No. 9, Article No. 1575, doi: 10.3390/app8091575.
[20] Keller, J.M., Gray, M.R., Givens, J.A. (1985). A fuzzy K-nearest neighbor algorithm, *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 15, No. 4, 580-585, doi: 10.1109/TSMC.1985.6313426.
[21] Gao, H.N., Shen, D.H., Yu, L., Zhang, W.C. (2020). Identification of cutting chatter through deep learning and classification, *International Journal of Simulation Modelling*, Vol. 19, No. 4, 667-677, doi: 10.2507/IJSIMM19-4-CO16.
[22] Wang, Y.D., Lu, X.C., Shen, J.R. (2021). Improved Genetic Algorithm (VNS-GA) using polar coordinate classification for workload balanced multiple Traveling Salesman Problem (mTSP), *Advances in Production Engineering & Management*, Vol. 16, No. 2, 173-184, doi: 10.14743/apem2021.2.392.
[23] Zhang, L., Ma, J., Liu, X., Zhang, M., Duan, X., Wang, Z. (2022). A novel support vector machine model of traffic state identification of urban expressway integrating parallel genetic and *c*-means clustering algorithm, *Tehnički Vjesnik – Technical Gazette,* Vol. 29, No. 3, 731-741, doi: 10.17559/TV-20211201014622.
[24] Tawhid, M.A., Savsani, V. (2019). Multi-objective sine-cosine algorithm (MO-SCA) for multi-objective engineering design problems, *Neural Computing and Applications*, Vol. 31, No. 2, 915-929, doi: 10.1007/s00521-017-3049-x.
[25] Chaouni Benabdellah, A., Benghabrit, A., Bouhaddou, I. (2019). A survey of clustering algorithms for an industrial context, *Procedia Computer Science,* Vol. 148, 291-302, doi: 10.1016/j.procs.2019.01.022.
[26] Gocken, T., Yaktubay, M. (2019). Comparison of different clustering algorithms via genetic algorithm for VRPTW, *International Journal of Simulation Modelling*, Vol. 18, No. 4, 574-585, doi: 10.2507/IJSIMM18(4)485.
[27] Kuo, R.J., Zheng, Y.R., Nguyen, T.P.Q. (2021). Metaheuristic-based possibilistic fuzzy k-modes algorithms for categorical data clustering, *Information Sciences*, Vol. 557, 1-15, doi: 10.1016/j.ins2020.12.051.
[28] Nguyen, T.P.Q., Kuo, R.J. (2019). Automatic fuzzy clustering using non-dominated sorting particle swarm optimization algorithm for categorical data, *IEEE Access*, Vol. 7, 99721-99734, doi: 10.1109/ACCESS.2019.2927593.
[29] Mirjalili, S. (2016). SCA: A sine cosine algorithm for solving optimization problems, *Knowledge-Based Systems*, Vol. 96, 120-133, doi: 10.1016/j.knosys.2015.12.022.
[30] Geng, Y.A., Li, Q., Zheng, R., Zhuang, F., He, R., Xiong, N. (2018). RECOME: A new density-based clustering algorithm using relative KNN kernel density, *Information Sciences*, Vol. 436-437, 13-30, doi: 10.1016/j.ins.2018.01.013.
[31] Cui, Y., Liu, H., Wang, Q., Zheng, Z., Wang, H., Yue, Z., Ming, Z., Wen, M., Feng, L., Yao, M. (2022). Investigation on the ignition delay prediction model of multi-component surrogates based on back propagation (BP) neural network, *Combustion and Flame*, Vol. 237, Article No. 111852, doi: 10.1016/j.combustflame.2021.111852.
[32] Dong, Y., Fu, Z., Peng, Y., Zheng, Y., Yan, H., Li, X. (2020). Precision fertilization method of field crops based on the Wavelet-BP neural network in China, *Journal of Cleaner Production*, Vol. 246, Article No. 118735, doi: 10.1016/j.jclepro.2019.118735.
[33] Sadowski, P. (2016). Notes on backpropagation, from http://web.khu.ac.kr/~tskim/MLPR%2025-3%20Back-propagation%20with%20CE%20(Handout).pdf, accessed 5 April 2023.
[34] Wang, K.-J., Hao, F.-J., Lee, Y.-X. (2022). A multiple-stage defect detection model by convolutional neural network, *Computers & Industrial Engineering*, Vol. 168, Article No. 108096, doi: 10.1016/j.cie.2022.108096.
[35] Chicco, D., Warrens, M.J., Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment, *IEEE Access*, Vol. 9, 78368-78381, doi: 10.1109/ACCESS.2021.3084050.

APEM journal
Advances in Production Engineering & Management
Volume 18 | Number 2 | June 2023 | pp 250–262
https://doi.org/10.14743/apem2023.2.471
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Simulation and Genetic Algorithm-based approach for multi-objective optimization of production planning: A case study in industry

Bojic, S.[a], Maslaric, M.[a,*], Mircetic, D.[b], Nikolicic, S.[a], Todorovic, V.[c]

[a]Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
[b]Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad, Serbia
[c]Science and Technology Park Novi Sad, Novi Sad, Serbia

A B S T R A C T

To stay competitive on the constantly changing and demanding market, production systems need to optimize their performance daily. This is particularly challenging in labour-intensive industries, which is characterized by highly volatile customer demand and significant daily variability of available workers. The Uncertainty related to the key production parameters in the industry is causing disruptions in long-term production planning and optimization, which leads to the long lead production times, operational risks and accumulation of inventory. To address these challenges, production systems need to ensure adequate operational production planning and optimization of all variables that are influencing the productivity of their systems on a daily basis. To tackle the problem, this study elaborates the application of discrete event simulations and genetic algorithm, using the Tecnomatix Plant Simulation software, to support decision-making and operational production planning and optimization in the industry. The simulation model developed for this purpose considers: customers demand changes, variable production times, operationally available resources and production batch size, to provide an optimal production sequence with the highest number of produced pieces and the lowest total work in process (WIP) inventory per day. To demonstrate the efficiency of the methodology and prove the benefits of the selected optimization approach, a case study is conducted in the textile factory.

A R T I C L E   I N F O

Keywords:
Discrete event simulation (DES);
Genetic algorithm (GA);
Production planning;
Multi-objective optimization;
Textile industry;
Tecnomatix Plant Simulation software

**Corresponding author:*
marinko@uns.ac.rs
(Maslaric, M.)

Article history:
Received 18 May 2023
Revised 8 June 2023
Accepted 28 June 2023

1. Introduction

To address the constantly growing customer demands the retailers put great pressure on the textile industry requesting frequent delivery of small quantities of a large number of different products [1, 2]. Satisfying such a stochastic demand with a high level of efficiency and low level of inventory requires effective operational planning and optimization of a production system, using new technologies that can support not only strategical, but also real-time, operational decision-making [3].

Many researchers investigated different influencing parameters which influence the efficiency of a production system, such as production process organization [4], product type [5], lot size scheduling [4] and operator's skill level [2]. However, there has been little research and even less practical implementation related to the relationship between these parameters and their common joint influence on the daily operational planning and optimization of a production system in the textile industry.

In today's industrial practice decision-making and operational planning still heavily rely on the production planners and their experience (expert knowledge) and future projections based on the data from the past. In those circumstances, without a proper digital tool, it is not possible to anticipate the impact of each influencing parameter and ensure optimal operational production planning and decision-making [6].

Considering the number, stochasticity and complexity of the influencing parameters, many researchers have used computer simulation, particularly discrete event simulation, to investigate the performance of production systems and to perform their optimization [2, 7-15].

High complexity of the optimization problem in production systems is also the reason why researchers are using genetic algorithms to reduce the computational time and ensure the quality of the obtained solution [16-18].

According to [19, 20], optimization solutions for production planning which satisfy the demand for increased volumes with an increased number of styles and personalized products, are a very actual topic in the textile industry. More specifically, over the past decades, several business and market trends have emerged that have reshaped the way the garment and textile industry is organized. This means that the textile manufacturers must be able to respond to a significant number of small and irregular orders, which would only be possible in the case of a more efficient and agile production system.

This demonstrates the significant need for practical simulation models that would simultaneously consider the impacts of all the typical parameters relevant for the daily operational production planning, while supporting real-time decision-making with the minimal required computational time.

This study presents an application of the discrete event simulations and genetic algorithm, using the Tecnomatix Plant Simulation software, to support decision-making and production planning and optimization in textile industry. The discrete event simulations enable complex systems representations including all relevant parameters, while the GA provides variation and sequencing of all influencing parameters in a short computational time required for the efficient operational planning. Although the approach is demonstrated in the textile industry, it can be equally efficiently used in other industry sectors. Of special interest are the labour intensive industries with the high variability of demand. To show the benefits of the proposed approach, a case study was analysed based on the validated simulation model of a textile factory. The simulation model considers following factors: the changing customers demand (variable in terms of different product types and quantity); the production times in relation to the product type and quantity; the operationally available resources and their distribution within the production system; as well as the production batch size; in order to provide the optimal production sequence which will result in the highest number of produced pieces per day with the lowest total work in process (WIP) inventory. In contrast to the papers using similar approach [19] for the strategical and tactical planning, this study investigates the benefits of the dynamic daily operational production planning, simultaneously considering the influence of all relevant variables. To fulfil this objective and contribute new knowledge, the paper is structured as follows. The next chapter provides a brief introduction into the research approach and literature review within the chosen topic. Chapter 3 formulates the case study and outlines simulation models for the company under consideration. Chapter 4 details the benefits of using genetic algorithm (GA) optimization in daily production planning, while in the fifth chapter a final discussion is presented, offering an overall summary and key findings.

2. Materials and methods

2.1 Research method

In this study we used literature review not only to provide an overview of the use of optimization tools in operational production planning, but also to highlight a specific type of industry in which additional research is needed. The literature review is followed by a case study as the research method suitable to understand how and why some phenomenon takes place in specific situations [21]. This review has shown that the textile industry still lacks sufficient research related to operational production planning optimization. Moreover, our case study showed that the textile industry certainly represents a type of industry in which the application of simulation models in operational production planning can contribute significant impact. A detailed analysis of the case study method can be found in [22]. Credibility of the case study is based on the transparency of the applied research, processes, and procedures, which allows verification by other researchers [23].

2.2 Literature review

Simulation is becoming one of the most used technologies to understand and analyse the dynamics of manufacturing systems [24]. It can be used as a descriptive tool to: realistically represent real-world systems, increase understanding of the relationships between the various components of a system; predict the performance of the system under new operating conditions; support the decision-making etc., without disrupting the ongoing production activities [25]. Simply, simulation is a problem-identification and solving tool that is flexible and less costly than physical prototyping and experimentation [26].

Computer simulation has been successfully used by many authors for planning and optimization of particular production lines or entire production and logistics systems in the textile industry. Simulation has been used for the shirt manufacturing production line design [27], for balancing the trousers production line [28], for balancing a sweat-shirt sewing line [29], for balancing the garment production line, in the five different scenarios, based on the production process times collected through RFID [10]. Some other authors used simulation for monitoring of the garment production line, comparing the production times gathered through the manual time taking and through the installed sensing [9]. The model-driven decision support system that is fed by real-time data, using data simulation and communication technologies was developed to improve the productivity of the manufacturing process of a garment manufacturing line [30].

Discrete event simulation (DES) model was proposed as the first step towards sustainable production scheduling in the textile manufacturing industry [6]. As DES is widely used to analyse activities of planning, implementation and operation of manufacturing and logistics systems, the commercially available DES software tools are commonly used for decision-making in different manufacturing systems [19]. Some studies that compared the different DES software tools recognized the Tecnomatix Plant Simulation by Siemens as the one with good visual aspects and the ability to integrate with other software [31]. Other studies have proven that Plant Simulation is a useful tool for optimal resource utilization [32].

Optimization problems which require minimization or maximization of functions with several variables and constraints can be very complex and difficult to solve with the use of conventional optimization methods. This is why they are commonly solved using algorithms based on the principle of evolution [2]. Some of the best-known meta-heuristic techniques, such as simulated annealing, the genetic algorithm (GA), and tabu search, were reviewed by some authors [33, 34] who concluded that these methods are remarkably effective in solving many types of optimization problems, particularly in finding near-optimal solutions for problems in complex multi-dimensional search spaces. Modified steady state GA was used for batch sizing and production scheduling in a hybrid flow shop with a limited buffer [35]. The genetic algorithm GA has been increasingly used for production optimization and operational management in the textile industry, such as loading, facility layout design, line balancing and lot sizing [2]. A genetic algorithm (GA) was successfully applied by different authors for optimizing the assignment of operatives in an assembly line [36] and for balancing workers' walking along the assembly line [37]. The unrelated parallel machines scheduling problem, in the textile industry, with machine and sequence-dependent setup times and limited resources, was addressed by applying the genetic algorithm GA [38]. Some researchers proposed a multi-objective GA model to schedule multiple products to different assembling stages considering the total costs and manufacturing capacities. Based on the experimental results they concluded that the GA is more powerful than any existing methods for solving production planning and scheduling problems [39]. Other authors combined simulation with the GA for stochastic lot size scheduling problems to define sequencing and lot size rules, which maximize the expected profit per time unit, demonstrating that GA simulation is the right approach to be used in complex environments [40].

3. Case study

To demonstrate the GA-supported DES approach to daily operational production planning, with short computation time, the simulation model for a textile company was developed using the Tecnomatix Plant Simulation software.

The model is developed and implemented in the company that produces high-quality shirts. The company was established over 20 years ago as a private family business and today is recognized as one of the best shirts manufacturing companies in the Republic of North Macedonia. The company produces more than 300,000 shirts annually and exports over 95 % of its production. The company's production system includes following four departments: (1) Tailoring and cutting; (2) Preparation 1 (collars and cuffs); (3) Preparation 2 (front, back and sleeves); (4) Shirt assembly and final control. Each of these departments has a different number of workers and performs a different list of operations. The sequence of operations, as well as the process times, depend on the model of the shirt. The material flow within the departments is organized with the conveyors ensuring the first in first out (FIFO) principle. The material flows between the last operation in one department and the first operation in the next department are organized with the production carts which also play the role of the WIP inventory holders between the departments. Production is realized in batches whereas the batch size depends on the model of the shirt and the customer order quantity, however, the most common batch size contains 10 product pieces.

3.1 Previous reality

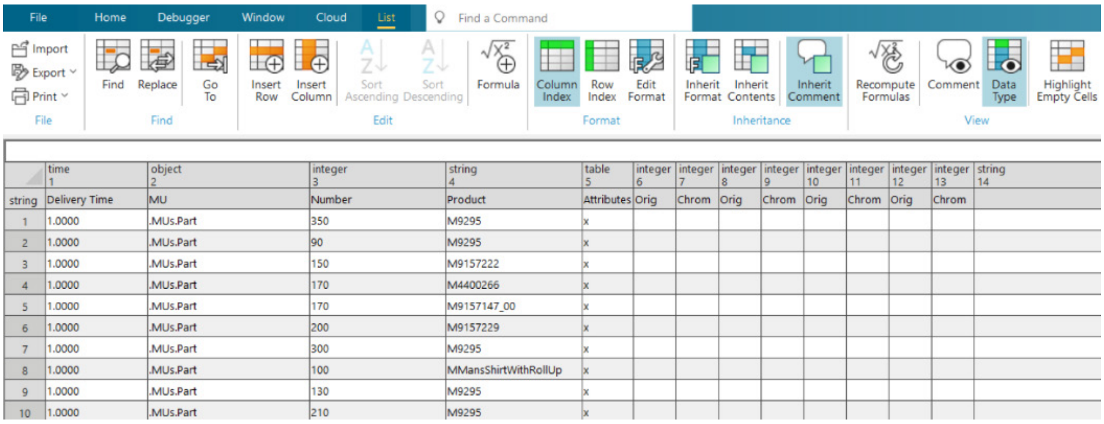
Before the simulation model was developed, production planning in the company was performed by the production manager, using Excel and the data on the customer orders from their Enterprise Resource System (ERP). Based on the customer orders and the operational lists for each model of the shirt, the production work orders were prepared and printed out to physically follow the material flows through the production system. The work orders include the data about the product (type, colour, size, etc.), batch size, list of operations that the product needs to go through, and process time for each of these operations. There is no real-time tracking of the products status or location through the system before the production is completed. The operational production scheduling in each department is performed by the department managers, based on their knowledge and experience. However, it is not coordinated among the departments, which slows the material flow and creates an uncontrolled WIP inventory accumulation.

The deep dive analysis of the production system revealed the following: unbalanced production between the departments; long production lead times, high WIP inventory, efficiency highly dependent on qualified labour availability and its distribution between the departments, no evidence or analysis on the influence of the production batch size and the production sequencing. The mentioned challenges are greatly conditioned with the daily variation of: product types and therewith the required production operations, demand-quantity of each product that needs to be produced and therewith the required production times per product, number of available workers in each department. Therefore, the simulation model was developed, as a supporting tool for the operational daily production planning that will ensure the highest possible production effectiveness while at the same time considering the influence of all relevant influential parameters, in a short computation time.

3.2 Way forward: Using simulation model

Modelling and simulation start with the proper identification of the problem and specification of simulation objectives. Once the objectives are set, it is required to determine which fixed system parameters and variables are relevant for the observed system [22].

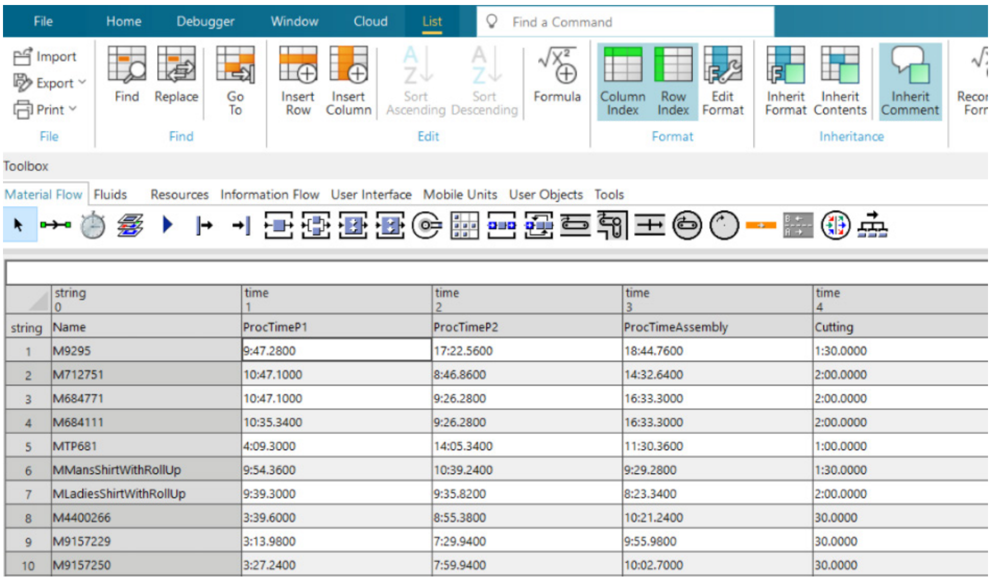
The customer demands (type and number of products, as well as the delivery times) represent the main input data for this simulation model. The data are gained from the real production system, for the period of ten weeks between June and August 2022. The company collects the customer orders on a weekly basis and sets the production plan based on the FIFO principle without the optimization of production sequencing according to the type or the quantity of the demanded products. The input of these data in the simulation was modelled through the delivery table called Production Plan (Fig. 1). In the table Production Plan, the first column defines the time when the products are entering the production system. Here, the times in all rows are set to one second, meaning that the products will enter the production system according to the PUSH principle and the FIFO sequencing. The customer demands, representing the number and the type of products that need to be produced, are defined in columns three and four, respectively. Special attributes of the products are defined in column five where the main attribute is the batch size of the product.



	time 1	object 2	integer 3	string 4	table 5	integer 6	integer 7	integer 8	integer 9	integer 10	integer 11	integer 12	integer 13	string 14
string	Delivery Time	MU	Number	Product	Attributes	Orig	Chrom	Orig	Chrom	Orig	Chrom	Orig	Chrom	
1	1.0000	.MUs.Part	350	M9295	x									
2	1.0000	.MUs.Part	90	M9295	x									
3	1.0000	.MUs.Part	150	M9157222	x									
4	1.0000	.MUs.Part	170	M4400266	x									
5	1.0000	.MUs.Part	170	M9157147_00	x									
6	1.0000	.MUs.Part	200	M9157229	x									
7	1.0000	.MUs.Part	300	M9295	x									
8	1.0000	.MUs.Part	100	MMansShirtWithRollUp	x									
9	1.0000	.MUs.Part	130	M9295	x									
10	1.0000	.MUs.Part	210	M9295	x									

Fig. 1 Production plan representing weekly customer demands

The production operations, including the real system process times, for every potential product type in each department, are defined in the table Production Times (Fig. 2). The production times are defined based on the real system measurements considering the type of the product, operations that the product is going through during production and process times of each operation. These production times are further recalculated in the model, in each simulation run, relative to the currently available number of workers in each department.



	string 0	time 1	time 2	time 3	time 4
string	Name	ProcTimeP1	ProcTimeP2	ProcTimeAssembly	Cutting
1	M9295	9:47.2800	17:22.5600	18:44.7600	1:30.0000
2	M712751	10:47.1000	8:46.8600	14:32.6400	2:00.0000
3	M684771	10:47.1000	9:26.2800	16:33.3000	2:00.0000
4	M684111	10:35.3400	9:26.2800	16:33.3000	2:00.0000
5	MTP681	4:09.3000	14:05.3400	11:30.3600	1:00.0000
6	MMansShirtWithRollUp	9:54.3600	10:39.2400	9:29.2800	1:30.0000
7	MLadiesShirtWithRollUp	9:39.3000	9:35.8200	8:23.3400	2:00.0000
8	M4400266	3:39.6000	8:55.3800	10:21.2400	30.0000
9	M9157229	3:13.9800	7:29.9400	9:55.9800	30.0000
10	M9157250	3:27.2400	7:59.9400	10:02.7000	30.0000

Fig. 2 Production times for each product in each department

The numbers of available operators in each department are defined as variables (*i, j, k, d*) which are replicated in the simulation model according to the real system, as well as the variable *s* – representing the total number of available workers in all departments. The initial number of workers in the department preparation 1, was set to I = 20, *the* number of workers in the department preparation 2, was set to j = 40, *the* number of workers in the assembly and final control department was set to k = 40, while the number of workers in the tailoring and cutting department was set to d = 4. Consequently, s was initially set to 104 and changed in some simulation runs according to the exact number of available workers on a particular day.

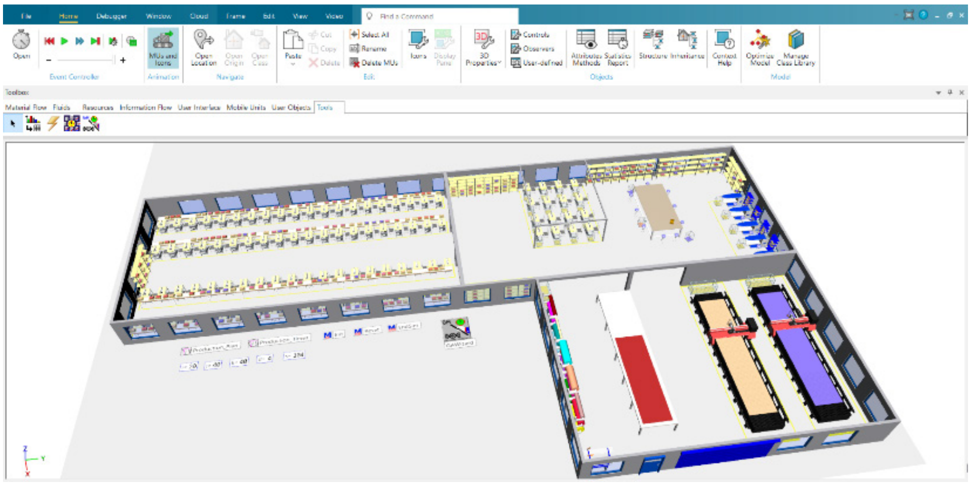


Fig. 3 Simulation model of the case study textile factory

The simulation model developed for this case study (Fig. 3) is based on the steps proposed by [21] and followed by other researchers [6, 15]. This approach ensures that the simulation model has all the necessary characteristics which allow it to be used as a decision-making tool.

In the first step, to ensure that the simulation model can be used as a production planning and decision-making tool, it was important to verify the modelling logic and validate the model results with the real production data. Verification, in this case, considered the use of low-speed graphic animation within the Tecnomatix Plant Simulation software, ensuring that the simulation model has the same logic of movement as the real system. Validation was performed by comparing the simulation results of the selected production operating days (the first day in each of the ten observed weeks) with the existing analytical data of the real system performance in those exact operating days. The simulation results of all simulation runs were within the 95 % confidence interval of the real system performance (Table 1) thus validating the accuracy and reliability of the simulation approach.

In the second step, the genetic algorithm was introduced in the model to perform optimization and increase production efficiency with adequate daily operational production planning and sequencing in comparison to the FIFO principle used in the real system.

3.3 New reality: Operational production planning and optimization with the simulation model using GA

Once the verification and validation of the simulation model were performed, a GA Wizard was implemented in the model, to provide optimization of the system operational production planning.

Two opposite objectives were set for the GA optimization:

- Maximize the number of produced products per day, considering the customer weekly demand and currently available number of workers, and
- Minimize the total WIP inventory, in order to reduce the product lead-time and to reduce the operational risks of having high inventory levels.

The optimization with the GA Wizard considers:

- Production sequence of work orders generated according to the customer demand, considering the weekly demand and daily available production resources;
- Distribution of available workers between the departments, considering the distribution limits that are defined based on the worker's skills;
- Optimal batch size, considering the defined limits that are defined based on the literature and the company's production planners' experience;
- Total WIP inventory level, considering the capacity limits that are defined based on the targeted inventory levels, to achieve the highest production efficiency and throughput with the lowest possible WIP inventory. It is hereby anticipated that the increased throughput in the system with the low WIP inventory will positively influence the product lead-time.

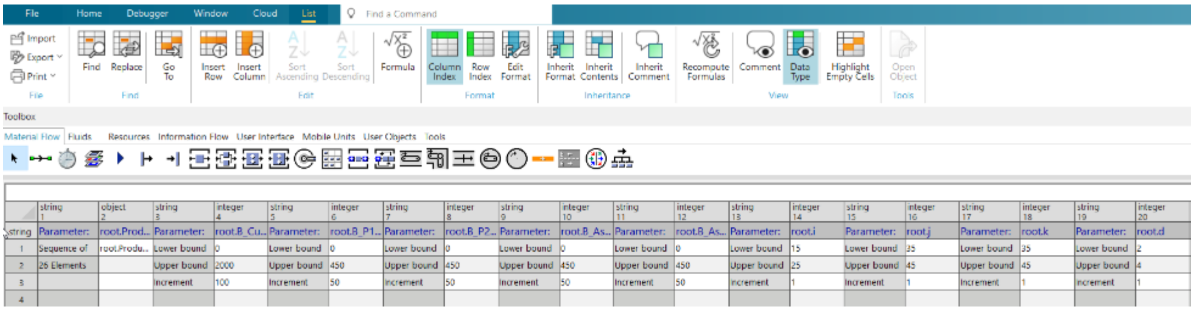


Fig. 4 Defined limits for the distribution of available workers and for the buffer capacities in each department

The total number of available workers in the real system is 104, however, it varies daily. Therefore, the variable *s* was initially set to 104, and adjusted in each simulation run to the number of workers that came to work on that exact date. The limits for the distribution of the available workers in each of the departments (variables *i, j, k, d*) were set to: $15 < i < 25$; $35 < j < 45$; $35 < k < 45$; $2 < d < 4$; with an increment of one (Fig. 4).

The limit for the batch size was set 1 < batch < 10, with an increment of one. The limit for the buffer after the department of tailoring and cutting was set to 2000 with an increment of 100, while for all other buffers between the other departments was set to 450, with an increment of 50 (Fig. 4).

The GA Wizard performed work orders (representing customer demands) sequencing while varying the number of workers in each department and varying the capacity of the available buffers, within the defined limits, to find the right sequence which will ensure the best fitness value considering the defined weightings of each of the defined objectives.

To model the opposing objectives, a variable called MinWIP was introduced and defined in an EndSim Method (Fig. 5). In the method, the MinWIP is calculated as the deviation between 10.000 and total WIP inventory between all production departments.

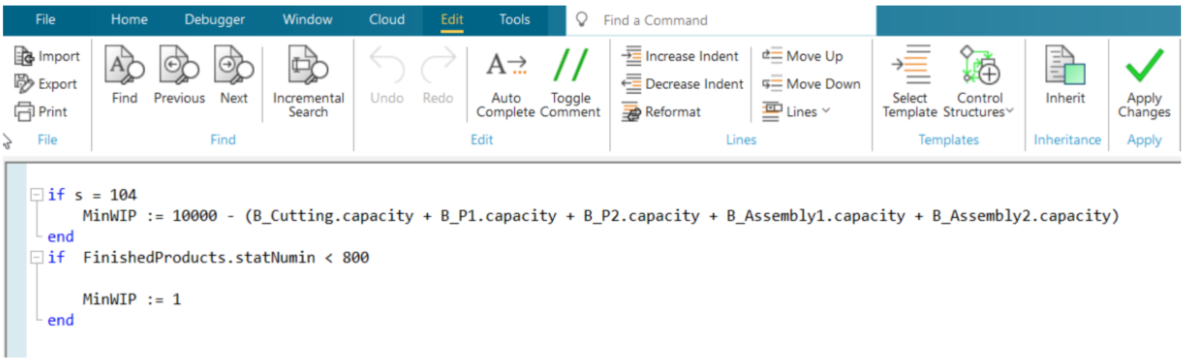


Fig. 5 EndSim Method, calculating the MinWIP and penalizing solutions that are not good enough to be considered

This formulation enables the minimization of the total WIP inventory to be reformulated into the maximization of the MinWIP variable (the higher the deviation, the lower the total WIP inventory) which enables the search for the optimal solution by maximizing both variables (the number of finished products and MinWIP) in the GA Wizard (Fig. 6). Additionally, the EndSim method is penalizing the solutions in which the number of produced products per day is less than 800, by simply setting the value of the MinWIP variable to one. In that way, the search for the optimal solution from the point of both objectives is facilitated.

The parameters used during the Tecnomatix GA Wizard optimization were: (i) one hundred generations; (ii) the size of generation is ten individuals; (iii) the elitist system was used for the selection, where the best solutions were used to generate offspring for the next generation; (iv) multi-objective optimization, and (v) weighting of the optimization objectives (Fig. 6).

The best optimization solution was searched as the best fines value expressed through the weighting of the two defined objectives, where the weights of the first and second objectives were based on the preferences of the company's production manager and were set to 0.7 and 0.3, respectively (Fig. 6).

Fig. 7 shows the evolution of optimized solutions against the number of generations, confirming that one hundred generations are sufficient for obtaining the optimization convergence in these simulation runs.

The developed model was tested for the daily operation of the first working day in each of the ten weeks. The main output of each simulation run consisted of the sequence of the production work orders (Fig. 8), distribution of workers to the production departments (Table 2) and the optimal batch size (Table 3).

Fig. 8 shows the GA optimization results (of the 5. simulation run): production buffers' required capacity, workers' distribution, batch size and production sequence.

This case study does not belong to computationally complex ones. Therefore, an average computation time of the ten simulation runs, for the defined parameters and limits, was 25 seconds. The computation time increases significantly with the problem complexity (in relation to the number and type of variables analysed in the system, their range and increment, as well as in relation to the number and size of generations of the GA). However, it can be expected that this approach can perform production optimisation and operational production planning of medium complex problems in just a few minutes.

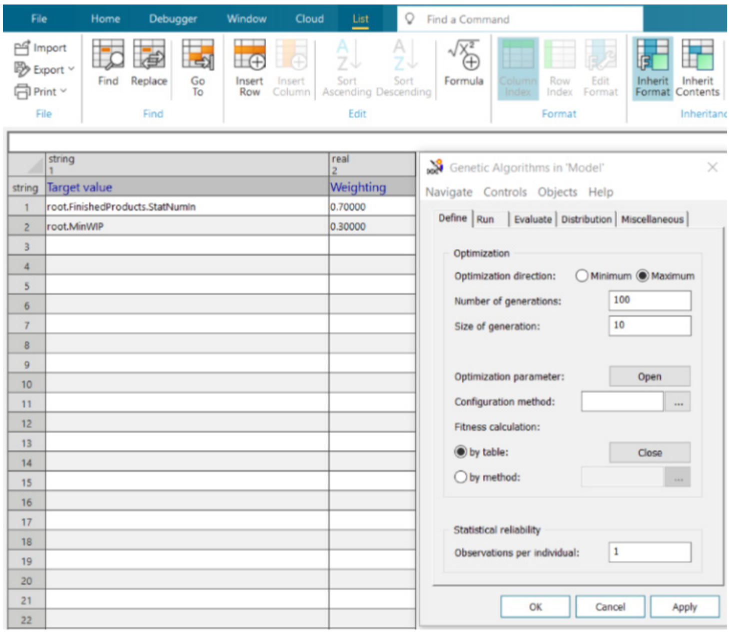


Fig. 6 Optimization parameters set in the GA Wizard of the Tecnomatix Plant Simulation

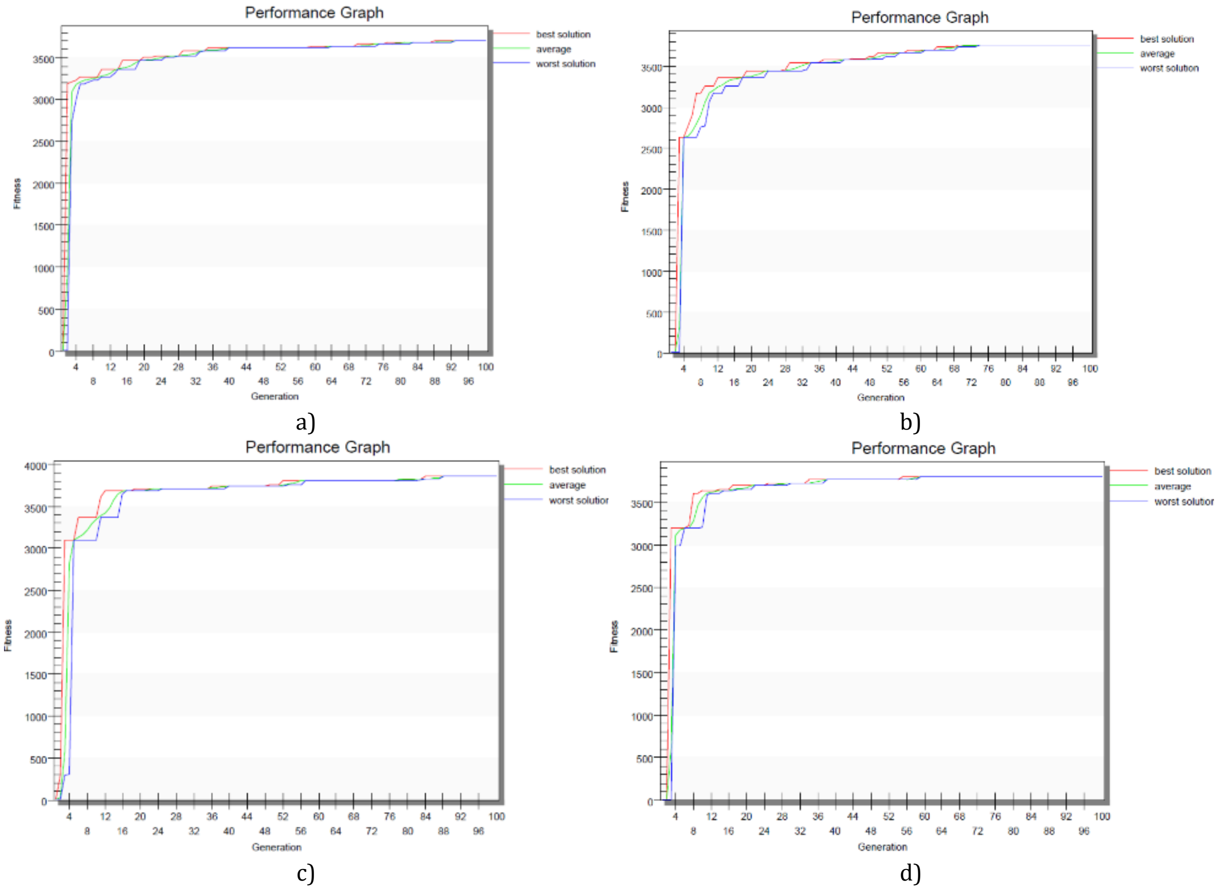


Fig. 7 Typical performance graphs of the ten performed simulation runs (a: 1, b: 5, c: 7, d: 10)

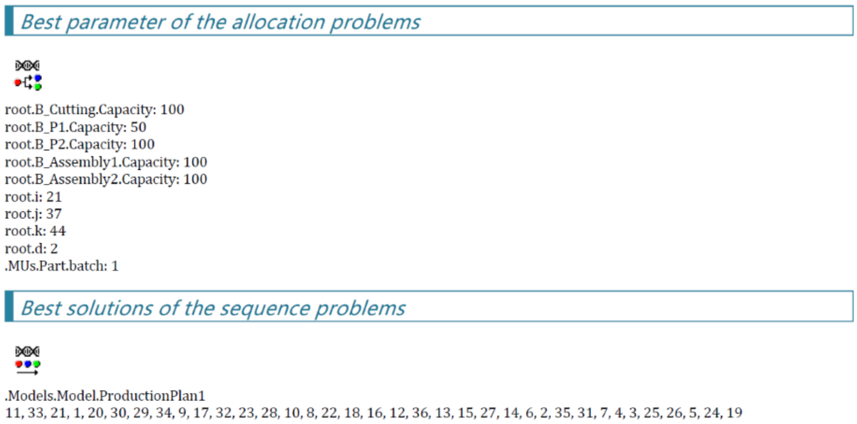


Fig. 8 GA optimization results, workers distribution, production buffers' capacity and production sequencing

4. Discussion

The simulation results of both developed models are presented in Table 1, where the basic one was used for the verification and validation of the simulation approach, while the second was integrated with the GA.

Comparison of the real system data and the simulation results without GA, undoubtedly verify the relevance and accuracy of the simulation model.

Comparison of the real system and the simulation results with the GA demonstrate great optimization potential of simulation and genetic algorithm for the operational production planning and sequencing (Fig. 9).

Table 1 Real system vs. simulation results – daily production and WIP in the selected days, in number of pieces

	Real system		Simulation		Simulation with GA	
Days	Production	WIP	Production	WIP	Production	WIP
1	1098	6654	1110	6588	1212	499
2	1000	4889	980	4958	1073	300
3	1153	8185	1130	8243	1310	391
4	1104	5154	1110	5128	1195	315
5	1245	5048	1260	4988	1268	434
6	1137	6253	1120	6348	1164	541
7	1117	5899	1100	5989	1160	409
8	1202	7742	1188	7758	1283	350
9	1031	6553	1020	6566	1204	537
10	1135	9235	1150	9099	1300	360

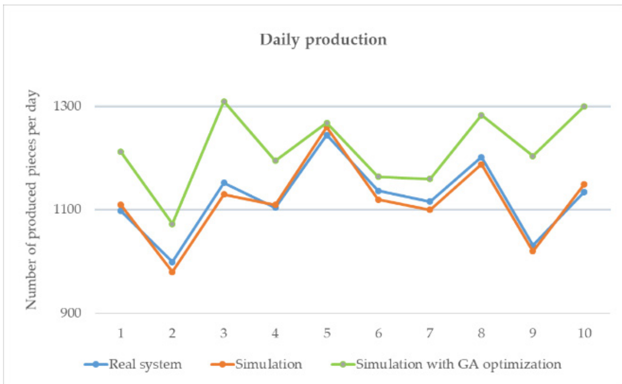


Fig. 9 Daily production – comparison of the real system analytics and simulation results

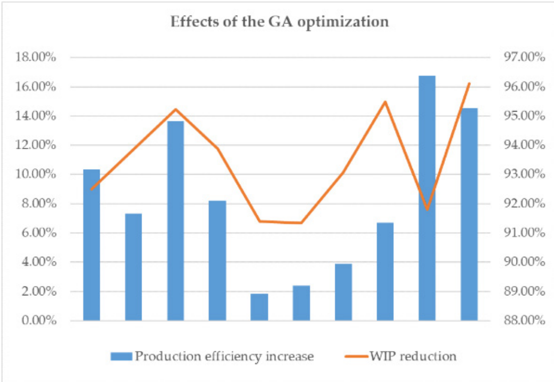


Fig. 10 Effects of the GA optimization

As shown in Fig. 10, the use of GA for the production sequencing can lead to an average increase in productivity between 8.5 % and 17 %. However, even more significant savings can be achieved through the reduction of WIP (Fig. 11).

According to the simulation results for the selected days, the operational production planning and sequencing with the use of GA and limitations of the WIP inventory, can lead to an average reduction of the total WIP between 93 % and 96 % (Fig. 10). This reduction is the result of the adequate production sequencing as well as the optimal distribution of the available workers between the departments and the optimization of batch size.

In the real system, the available workers were always distributed in the same way. Contrastingly, in the simulation model, the GA optimization proposed different workers' distribution between the departments in each simulation run (Table 2).

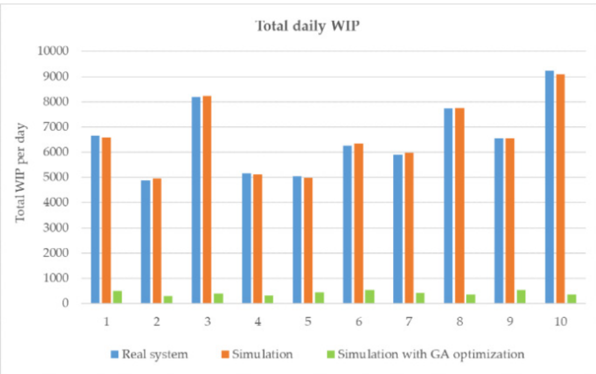


Fig. 11 Total WIP – comparison of the real system analytics and simulation results

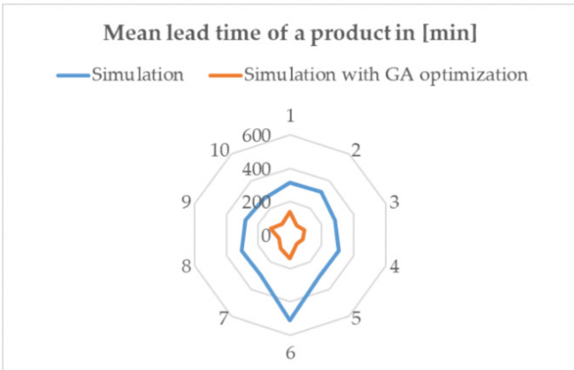


Fig. 12 Comparison of the mean production lead times for the simulated products (min)

Table 2 Distribution of available workers between the departments in real system and in each simulation run

	Real system					Simulation with GA			
	s	*i*	*j*	*k*	*d*	*i*	*j*	*k*	*d*
1	104	20	40	40	4	22	37	43	2
2	96	18	36	40	2	20	37	36	3
3	104	20	40	40	4	22	39	39	4
4	104	20	40	40	4	25	40	36	3
5	104	20	40	40	4	21	37	44	2
6	99	20	38	37	4	21	37	39	2
7	95	17	39	35	4	22	35	36	2
8	104	20	40	40	4	24	38	39	3
9	104	20	40	40	4	24	38	39	3
10	104	20	40	40	4	25	35	41	3

Table 3 GA optimization of batch size and its influence on assortment size

	Real system		GA optimization	
Day	Batch size	Assortment size	Batch size	Assortment size
1	10	5	1	6
2	10	3	1	6
3	10	5	6	4
4	10	5	5	6
5	10	7	1	7
6	10	7	1	6
7	10	1	1	4
8	10	4	1	6
9	10	5	3	5
10	10	3	10	4

Table 3 shows GA optimization of production batch size in comparison to the constant batch size that is determined in the real system. The simulation results of the assortment size (Table 3) prove that the increase of productivity through the GA optimization is not due to the assortment size reduction, as one might suppose, but a consequence of the optimal sequencing, as well as the resource and material flow distribution.

Furthermore, the simulation results proved that such an extreme reduction of the WIP and optimization of the batch size, as expected, positively influenced the product lead time. Based on the simulation results, the reduction of product lead time was more than 3 times (from average 318 min to average 95 min). The mean value of the lead time reduction, per day for all products, was between 70 % and 78 % (Fig. 12).

These results prove and verify the effectiveness of the DES and GA approach for operational production planning and optimization in the case study.

5. Conclusions

The textile industry is challenged by highly stochastic and unpredictable demand, as well as specifically stochastic production performance caused by the predominantly manual operation and low level of digitalization, which are resulting in high levels of WIP inventory and long product lead-time. At the same time, the production planning and resource utilization optimization is based on outdated data and the experience of the production planners. This is not specific to just the textile industry, but is common in other industries as well, particularly the labour intensive ones with the low level of automation and digitalization. The approach proposed within this case study investigates the possibility to improve production performance in industry by improving operational production planning and optimization. The approach considers the application of discrete event simulation and genetic algorithm, to perform the operational production sequencing of the customer orders, distribute the available workers to the work departments, and determine the optimal batch size, in a way to maximize productivity and minimize the WIP inventory and product lead time. The proposed approach was tested using the Tecnomatix Plant Simulation software to eliminate the weaknesses of the manual scheduling procedures based on the judgment and experience of the production planners. Although it was applied to the case study of the shirts production company, the proposed approach is not limited to the textile industry but could be efficiently applied in other industry sectors. The results of this case study validated the ascendency of the simulation approach in operational planning and optimization, as well as the high efficiency of the genetic algorithm. Future steps in the research and practice should consider integration of the simulation software with the existing ERP system to ensure real-time data consideration, as well as with the potential MES (Manufacturing Execution System) which would transform the simulation model into the production system digital shadow. In that way, the simulation model would become a real-time dynamic operational planning tool with the capacity to perform not only the daily but also the real-time production scheduling and optimization.

Acknowledgement

The realization of this paper has been supported by the Serbian Ministry of Education, Science and Technological Development program through the projects no. 451-03-68/2020-14/200156: "Innovative scientific and artistic research from the FTS (activity) domain", by the Danube Transnational Programme, through the project DTP1-050-3.1: "Regional and Transport Development in the Danube-Black Sea Region towards the Transnational Multiport Gateway Region (DBS Gateway Region)", as well as by the FP7 Programme, through the project no. 314005: "Development of a Next generation European Inland Waterway Ship and Logistics System (NEWS)".

References

[1] Bhardwaj, V., Fairhurst, A. (2010). Fast fashion: Response to changes in the fashion industry, *The International Review of Retail, Distribution and Consumer Research*, Vol. 20, No. 1, 165-173, doi: 10.1080/09593960903498300.

[2] Mak, L.C.L. (2016). *Modelling of apparel production system and optimization of lot size scheduling in apparel industry,* PhD. thesis, Hong Kong Polytechnic University, Hong Kong.

[3] Ahmad, S., Miskon, S., Alabdan, R., Tlili, I. (2020). Towards sustainable textile and apparel industry: Exploring the role of business intelligence systems in the era of Industry 4.0, *Sustainability*, Vol. 12, No. 7, Article No. 2632, doi: 10.3390/su12072632.

[4] Ekren, B.Y., Ornek, A.M. (2008). A simulation based experimental design to analyze factors affecting production flow time, *Simulation Modelling Practice and Theory*, Vol. 16, No. 3, 278-293, doi: 10.1016/j.simpat.2007.11.016.

[5] Kincade, D.H., Kim, J., Kanakadurga, K.S. (2013). An empirical investigation of apparel production systems and product line groups through the use of collar designs, *Journal of Textile and Apparel, Technology and Management*, Vol. 8, No. 1, 1-15.

[6] Pirola, F., Zambetti, M., Cimini, C. (2021). Applying simulation for sustainable production scheduling: A case study in the textile industry, *IFAC-PapersOnLine*, Vol. 54, No. 1, 373-378, doi: 10.1016/j.ifacol.2021.08.041.

[7] Ramesh, V., Sreenivasa Prasad, K., Srinivas, T. (2009). Study on implementation of one-piece lean line design using simulation techniques: A practical approach, *Journal of Industrial Engineering International*, Vol. 5, No. 8, 20-36.

[8] Brahmadeep, Thomassey, S. (2014). A simulation based comparison: Manual and automatic distribution setup in a textile yarn-rewinding unit of a yarn-dyeing factory, *Simulation Modelling Practice and Theory*, Vol. 45, 80-90, doi: 10.1016/j.simpat.2014.04.002.

[9] Wang, C.-N., Cheng, Z.-H., Phan, N.K.P., Nguyen, V.T. (2021). Scheduling optimization modelling: A case study of a woven label manufacturing company, *Computer Systems Science and Engineering*, Vol. 38, No. 2, 239-249, doi: 10.32604/csse.2021.016578.

[10] Jung, W.-K., Kim, H., Park, Y.-C., Lee, J.-W., Suh, E.S. (2020). Real-time data-driven discrete-event simulation for garment production lines, *Production Planning & Control*, Vol. 33, No. 5, 480-491, doi: 10.1080/09537287.2020.1830194.

[11] Sime, H., Jana, P., Panghal, D. (2019). Feasibility of using simulation technique for line balancing in apparel industry, *Procedia Manufacturing,* Vol. 30, 300-307, doi: 10.1016/j.promfg.2019.02.043.

[12] Grznar, P., Gregor, M., Gola, A., Nielsen, I., Mozol, S., Seliga, V. (2022). Quick workplace analysis using simulation, *International Journal of Simulation Modelling*, Vol. 21, No. 3, 465-476, doi: 10.2507/IJSIMM21-3-612.

[13] Ištoković, D., Perinić, M., Borić, A. (2021). Determining the minimum waiting times in a hybrid flow shop using simulation-optimization approach, *Tehnički Vjesnik – Technical Gazette,* Vol. 28, No. 2, 568-575, doi: 10.17559/TV-20210216132702.

[14] Li, G.Z., Xu, Z.G., Yang, S.L., Wang, H.Y., Bai, X.L., Ren, Z.H. (2020). Bottleneck identification and alleviation in a blocked serial production line with discrete event simulation: A case study, *Advances in Production Engineering & Management,* Vol. 15, No. 2, 125-136, doi: 10.14743/apem2020.2.353.

[15] Yang, S.L., Wang, J.Y., Xin, L.M., Xu, Z.G. (2022). Verification of intelligent scheduling based on deep reinforcement learning for distributed workshops via discrete event simulation, *Advances in Production Engineering & Management,* Vol. 17, No. 4, 401-412, doi: 10.14743/apem2022.4.444.

[16] Man, K.F., Tang, K.S., Kwong, S., Ip, W.H. (2000). Genetic algorithm to production planning and scheduling problems for manufacturing systems, *Production Planning & Control,* Vol. 11, No. 5, 443-458, doi: 10.1080/ 09537280050051942.

[17] Alhijawi, B., Awajan, A. (2023). Genetic algorithms: Theory, genetic operators, solutions, and applications, *Evolutionary Intelligence*, Vol. 16, doi: 10.1007/s12065-023-00822-6.

[18] Lee, C.K.H. (2018). A review of applications of genetic algorithms in operations management, *Engineering Applications of Artificial Intelligence*, Vol. 76, 1-12, doi: 10.1016/j.engappai.2018.08.011.

[19] Ferro, R., Cordeiro, G.A., Ordóñez, R.E.C., Beydoun, G., Shukla, N. (2021). An optimization tool for production planning: A case study in a textile industry, *Applied Science*, Vol. 11, No. 18, Article No. 8312, doi: 10.3390/ app11188312.

[20] Parven, I., Mia, M.A., Ali, M.S., Rafsun-Ul-Hasan, K., Rahman, M.M., Mahmud, M.I., Cho, H.M. (2019). Implementation of JIT to increase productivity in sewing section of garment industry, In: *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Bangkok, Thailand, 618-629.

[21] Yin, R.K. (2010). Case study research: Design and methods, 4th edition, Sage, Los Angeles, USA.

[22] Elgazzar, S. (2013). *Enhancing the company's financial performance through the performance of supply chain operations: A case study of an Egyptian manufacturing company,* Doctoral thesis, University of Huddersfield, Huddersfield, United Kingdom.

[23] Ellram, L.M. (1996). The use of the case study method in logistics research, *Journal of Business Logistics*, Vol. 17, No. 2, 93-138.

[24] Jeon, S.M., Kim, G. (2016). A survey of simulation modelling techniques in production planning and control (PPC), *Production Planning & Control,* Vol. 27, No. 5, 360-377, doi: 10.1080/09537287.2015.1128010.

[25] Law, A.M., Kelton, W.D. (1991). *Simulation modelling and analysis*, 2nd edition, McGraw-Hill, New York, USA.

[26] Nyemba, W.R., Mbohwa, C. (2017). Modelling, simulation and optimization of the materials flow of a multi-product assembling plant, *Procedia Manufacturing*, Vol. 8, 59-66, doi: 10.1016/j.promfg.2017.02.007.

[27] Güner, M.G., Ünal, C. (2008). Line balancing in the apparel industry using simulation techniques, *Fibres & Textiles in Eastern Europe*, Vol. 16, No. 2, 75-78.

[28] Fozzard, G., Spragg, J., Tyler, D. (1996). Simulation of flow lines in clothing manufacture, Part 1: Model construction, *International Journal of Clothing Science and Technology*, Vol. 8, No. 4, 17-27, doi: 10.1108/09556229610123982.

[29] Kursun, S., Kalaoglu, F. (2009). Simulation of production line balancing in apparel manufacturing, *Fibres & Textiles in Eastern Europe*, Vol. 17, No. 4, 68-71.

[30] Ivatury, V.M.K., Bonsa, K.B. (2022). Productivity improvement of a garment assembly line using simulation and model driven decision support system, *ARPN Journal of Engineering and Applied Sciences*, Vol. 17, No. 4, 407-415.

[31] Rashidi, Z., Rashidi, Z. (2016). Evaluation and ranking of discrete simulation tools, *Journal of Electrical and Computer Engineering Innovations*, Vol. 4, No. 1, 69-84, doi: 10.22061/jecei.2016.562.

[32] Butrat, A., Supsomboon, S. (2022). A Plant Simulation approach for optimal resource utilization: A case study in the tire manufacturing industry, *Advances in Production Engineering & Management,* Vol. 17, No. 2, 243-255, doi: 10.14743/apem2022.2.434.

[33] Silberholz, J., Golden, B. (2010). Comparison of metaheuristics, In: Gendreau, M., Potvin, J.Y. (eds.), *Handbook of metaheuristics, International series in operations research & Management science,* Vol. 146, Springer, Boston, USA, 625-640, doi: 10.1007/978-1-4419-1665-5_21.

[34] Sivanandam, S., Deepa, S. (2008). Genetic algorithm optimization problems, In: *Introduction to genetic algorithms,* Springer, Berlin, Germany, 165-209, doi: 10.1007/978-3-540-73190-0_7.

[35] Janeš, G., Ištokovic, D., Jurković, Z, Perinić, M. (2022). Application of modified steady-state genetic algorithm for batch sizing and scheduling problem with limited buffers, *Applied Science*, Vol. 12, No. 22, Article No. 11512, doi: 10.3390/app122211512.

[36] Wong, W.K., Guo, Z.X., Leung, S.Y.S. (2013). *Optimizing decision making in the apparel supply chain using artificial intelligence (AI): From production to retail (Woodhead publishing series in textiles),* 1st edition, Woodhead Publishing Limited, Cambridge, USA, doi: 10.1533/9780857097842.

[37] Al-Zuheri, A., Luong, L., Xing, L. (2016). Developing a multi-objective genetic optimisation approach for an operational design of a manual mixed-model assembly line with walking workers, *Journal of Intelligent Manufacturing*, Vol. 27, 1049-1065, doi: 10.1007/s10845-014-0934-3.

[38] Berthier, A., Yalaoui, A., Chehade, H., Yalaoui, F., Amodeo, L., Bouillot, C. (2022). Unrelated parallel machines scheduling with dependent setup times in textile industry, *Computers & Industrial Engineering*, Vol. 174, Article No. 108736, doi: 10.1016/j.cie.2022.108736.

[39] Man, K.F., Tang, K.S., Kwong, S., Ip, W.H. (2020). Genetic algorithm to production planning and scheduling problems for manufacturing systems, *Production Planning & Control*, Vol. 11, No. 5, 443-458, doi: 10.1080/09537280050051942.

[40] Kämpf, M., Köchel, P. (2006). Simulation-based sequencing and lot size optimisation for a production-and-inventory system with multiple items, *International Journal of Production Economics*, Vol. 104, No. 1, 191-200, doi: 10.1016/j.ijpe.2006.02.008.

Calendar of events

- European Simulation and Modelling Conference (ESM 2023), October 24-26, 2023, Toulouse, France.
- 17th International Conference on Industrial and Manufacturing Systems Engineering, November 27-28, 2023, London, United Kingdom.
- 18th International Conference on Advanced Manufacturing Engineering and Technologies,

  January 15-16, 2024, Montevideo, Uruguay.
- 2024 Annual Modeling and Simulation Conference (ANNSIM 2024), May 20-23, 2024, Washington D.C., USA.
- 18th International Conference on Industrial and Manufacturing Systems Engineering, August 9-10, 2024, Lagos, Nigeria.

This page intentionally left blank.

Notes for contributors

General

Articles submitted to the *APEM journal* should be original and unpublished contributions and should not be under consideration for any other publication at the same time. Manuscript should be written in English. Responsibility for the contents of the paper rests upon the authors and not upon the editors or the publisher. The content from published paper in the *APEM journal* may be used under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0). For most up-to-date information please see the APEM journal homepage *apem-journal.org*.

Submission of papers

A submission must include the corresponding author's complete name, affiliation, address, phone and fax numbers, and e-mail address. All papers for consideration by *Advances in Production Engineering & Management* should be submitted by e-mail to the journal Editor-in-Chief:

Miran Brezocnik, Editor-in-Chief
UNIVERSITY OF MARIBOR
Faculty of Mechanical Engineering
Chair of Production Engineering
Smetanova ulica 17, SI – 2000 Maribor
Slovenia, European Union
E-mail: editor@apem-journal.org

Manuscript preparation

Manuscript should be prepared in *Microsoft Word 2010* (or higher version) word processor. *Word .docx* format is required. Papers on A4 format, single-spaced, typed in one column, using body text font size of 11 pt, should not exceed 12 pages, including abstract, keywords, body text, figures, tables, acknowledgements (if any), references, and appendices (if any). The title of the paper, authors' names, affiliations and headings of the body text should be in *Calibri* font. Body text, figures and tables captions have to be written in *Cambria* font. Mathematical equations and expressions must be set in *Microsoft Word Equation Editor* and written in *Cambria Math* font. For detail instructions on manuscript preparation please see instruction for authors in the *APEM journal* homepage *apem-journal.org*.

The review process

Every manuscript submitted for possible publication in the *APEM journal* is first briefly reviewed by the editor for general suitability for the journal. Notification of successful submission is sent. After initial screening, and checking by a special plagiarism detection tool, the manuscript is passed on to at least two referees. A double-blind peer review process ensures the content's validity and relevance. Optionally, authors are invited to suggest up to three well-respected experts in the field discussed in the article who might act as reviewers. The review process can take up to eight weeks on average. Based on the comments of the referees, the editor will take a decision about the paper. The following decisions can be made: accepting the paper, reconsidering the paper after changes, or rejecting the paper. Accepted papers may not be offered elsewhere for publication. The editor may, in some circumstances, vary this process at his discretion.

Proofs

Proofs will be sent to the corresponding author and should be returned within 3 days of receipt. Corrections should be restricted to typesetting errors and minor changes.

Offprints

An e-offprint, i.e., a PDF version of the published article, will be sent by e-mail to the corresponding author. Additionally, one complete copy of the journal will be sent free of charge to the corresponding author of the published article.

APEM journal

Advances in Production Engineering & Management

Chair of Production Engineering (CPE)
University of Maribor
APEM homepage: apem-journal.org

Volume 18 | Number 2 | June 2023 | pp 133-266

Contents

Scope and topics 136

Real-time scheduling for dynamic workshops with random new job insertions by using deep reinforcement learning 137
Sun, Z.Y.; Han, W.M.; Gao, L.L.

Genetic algorithm-based approach for makespan minimization in a flow shop with queue time limits and skip-ping jobs 152
Han, J.H.; Lee, J.Y.

Hybrid forecasting modelling of cost and time entities for planning and optimizing projects in the die-cast aluminium industry 163
Muñoz-Ibañez, C.; Chairez, I.; Jimenez-Martinez, M.; Molina, A.; Alfaro-Ponce, M.

When core sorting and quality grading is beneficial to remanufacturers: Insights from analytical models 175
Cao, X.H.; Shi, X.L.; Lan, H.J.; Huang, D.

Engineering-to-order manufacturing: A criticality analysis of key challenges and solutions based on literature review 187
Fortes, C.S.; Tenera, A.B.; Cunha, P.F.; Teixeira, J.P.

Ranking dominant losses in small and medium-sized enterprises (SMEs) in the context of the lean concept application 199
Kondić, V.; Maglić, L.; Runje, L.; Marić, D.

An improved discrete particle swarm optimization approach for a multi-objective optimization model of an urban logistics distribution network considering traffic congestion 211
Li, K.; Li, D.; Ma, H.Q.

Blockchain-based tripartite evolutionary game study of manufacturing capacity sharing 225
Wang, T.Y.; Zhang, H.

Enhancing automated defect detection through sequential clustering and classification: An industrial case study using the Sine-Cosine Algorithm, Possibilistic Fuzzy c-means, and Artificial Neural Network 237
Nguyen, T.P.Q.; Yang, C.L.; Le, M.D.; Nguyen, T.T.; Luu, M.T.

Simulation and Genetic Algorithm-based approach for multi-objective optimization of production planning: A case study in industry 250
Bojic, S.; Maslaric, M.; Mircetic, D.; Nikolicic, S.; Todorovic, V.

Calendar of events 263

Notes for contributors 265

Published by CPE, University of Maribor

ISSN 1854-6250

9 771854 625008

apem-journal.org