ISSN 1854-6250

APEM journal

Advances in Production Engineering & Management

Volume 16 | Number 3 | September 2021

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

Subsidizer: The journal is subsidized by Slovenian Research Agency

Published by CPE, University of Maribor.

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 16 | Number 3 | September 2021 | pp 265–388

Contents

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
Volume 16 | Number 3 | September 2021 | pp 269–284
https://doi.org/10.14743/apem2021.3.399

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

A new solution to distributed permutation flow shop scheduling problem based on NASH Q-Learning

Ren, J.F.[a,b], **Ye, C.M.**[a,*], **Li, Y.**[a]

[a]School of Business, University of Shanghai for Science and Technology, Shanghai, P.R. China
[b]School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou, P.R. China

A B S T R A C T

Aiming at Distributed Permutation Flow-shop Scheduling Problems (DPFSPs), this study took the minimization of the maximum completion time of the workpieces to be processed in all production tasks as the goal, and took the multi-agent Reinforcement Learning (RL) method as the main frame of the solution model, then, combining with the NASH equilibrium theory and the RL method, it proposed a NASH Q-Learning algorithm for Distributed Flow-shop Scheduling Problem (DFSP) based on Mean Field (MF). In the RL part, this study designed a two-layer online learning mode in which the sample collection and the training improvement proceed alternately, the outer layer collects samples, when the collected samples meet the requirement of batch size, it enters to the inner layer loop, which uses the Q-learning model-free batch processing mode to proceed and adopts neural network to approximate the value function to adapt to large-scale problems. By comparing the Average Relative Percentage Deviation (ARPD) index of the benchmark test questions, the calculation results of the proposed algorithm outperformed other similar algorithms, which proved the feasibility and efficiency of the proposed algorithm.

A R T I C L E   I N F O

Keywords:
Flow shop scheduling;
Distributed scheduling;
Permutation flow shop;
Reinforcement learning;
NASH Q-learning;
Mean field (MF)

**Corresponding author:*
171910083@st.usst.edu.cn
(Ye, C.M.)

Article history:
Received 29 July 2021
Revised 8 September 2021
Accepted 26 September 2021

1. Introduction

Under intelligent manufacturing mode, cross-product value chain integration has become a normal situation, for example, many parts of air conditioners and refrigerators are similar or even the same, so to reduce costs, they can be produced and processed via distributed production. On this basis, this paper abstracted the problem of distributed production in different geographical locations to a DFSP, and divided it into two sub-problems, the reasonable allocation of workpieces among factories, and the processing sequence arrangement within a factory.

Field scholars have achieved various research results in terms of DFSP [1-3], for example, Wang *et al.* [4] analyzed the research status and development direction of DFSPs. Fernandez-Viagas and Framinan [5] proposed a new heuristic algorithm and obtained better upper boundaries. In terms of DPFSP, Gao *et al.* [6, 7] used heuristic algorithm to solve the problem, Naderi and Ruiz [8] proposed a scatter search algorithm, Lin *et al.* [9] proposed an improved iterative greedy algorithm, Fernandez-Viagas and Framinan [5] proposed a bounded search iterative greedy algorithm, Yang *et al.* [10] comprehensively optimized the assembly transport strategy, production process, and production configuration of a reconfigurable flow shop, Wang *et al.* [11] proposed a distributed estimation algorithm, and they used their respective algorithm to solve the problem. Rifai *et al.* [12] proposed a multi-objective adaptive large neighborhood search algorithm based on the Pareto front to study the distributed reentrant arrangement flow shop. Wang *et al.* [13] proposed a hybrid distributed estimation algorithm based on fuzzy logic to solve the production scheduling problem with the maximum completion time criterion under machine failure conditions. Komaki and Malakooti [14] proposed a generalized variable neighborhood search metaheuristic algorithm to solve the problem of minimizing the maximum completion time of the distributed no-waiting flow shop scheduling problem. Chen *et al.* [15] proposed the non-dominated sorting genetic algorithm (NSGA) to the design of non-compact flow shop scheduling plan, and successfully solved the multi-objective optimization problem considering process connection, Lebbar *et al.* [16] proposed a computational evaluation of a new mixed integer linear programming (MILP) model developed for the multi-machine flowshop scheduling, Rathinam *et al.* [17] proposed heuristic based methodology to solve permutation flow shop scheduling.

As the multi-agent Deep Reinforcement Learning (DRL) technology [18] is developing rapidly, the multi-agent application of the Q-learning algorithm has become more prominent [19-22], and scholars have successively applied this new technology to combinatorial optimization problems such as workshop scheduling, and obtained a series of research results. For example, a few studies [23-26] explored deep into the communication and cooperation between the multiple agents of DRL. Omidshafiei *et al.* [27] studied multi-task and multi-agent RL problem, and proposed a method to upgrade single-task strategy to multi-task strategy. Sunehag *et al.* [28] used a new value decomposition network structure to train individual agents to solve the joint reward problem of multiple agents. Li *et al.* [29] proposed a workflow job scheduling algorithm based on load balancing, and the research results showed that the proposed algorithm can effectively utilize the cloud resources. Neary *et al.* [30] studied the interaction of collaborative multi-agent RL in a shared environment. Some scholars applied Nash equilibrium and multi-agent RL algorithm to optimal control problems [31, 32]. Shah *et al.* [33] studied the Nash equilibrium RL problem of the two-person zero-sum game. Feng *et al.* [34] proposed a polynomial time algorithm for dual-agent scheduling problem to find all Pareto optimal solutions.

In summary, this paper took DFSP as the research object and proposed new NASH-Q DRL algorithm to solve DFSP based on existing multi-agent RL theory and algorithm results, and the problem was solved by a data-driven method, which avoided the design of complex heuristic rules.

2. Problem description

Distributed workshop scheduling can be divided into several types: distributed parallel machine scheduling, distributed flow shop scheduling, distributed job shop scheduling, distributed assembly scheduling, and distributed flexible workshop scheduling. This paper aims at the DPFSP, namely the distributed permutation flow shop scheduling problem, before solving the problem, it's assumed that each factory has one flow production line, the production capacity of each factory is the same, each factory has the same workshop layout, the same number of machines, and the same processing ability for all workpieces.

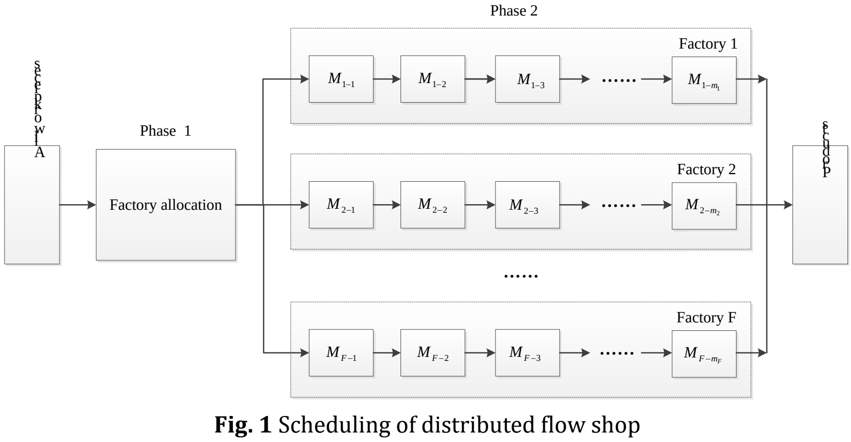


Fig. 1 Scheduling of distributed flow shop

In DPFSP, the first thing to do is to solve the workpiece allocation problem in each workshop, that is, to allocate workpieces to be processed to each workshop according to certain initial conditions; the second is to solve the workpiece processing sequence on the production line in each workshop, that is, to arrange the processing sequence of the workpieces that have been allocated to a same workshop, and use machines to process the sorted workpieces one by one. Due to mutual coupling, the problem shows high complexity, Fig. 1 gives a diagram of the scheduling of distributed flow shop.

In problem solving, the goal is to minimize the maximum completion time of all workpieces. The maximum completion time of DPFSP means the total completion time of all the workpieces, which is determined by the workshop with the longest completion time, and this workshop is called the key workshop in this study. Assume: a processing task has a total of n workpieces that need to be processed in F workshops, each workshop has m machines, and each workpiece needs to go through m processing procedures, the processing procedures of all workpieces are the same, once the workshop is decided for a workpiece, all procedures need to be completed within this workshop, the numbers of workpieces allocated to each workpiece are $(n_1, n_2, \dots, n_f)$, the processing time of workpiece i on machine j in workshop f is denoted as $p_{i,j}^f$, the completion time of the k-th workpiece on machine j is denoted as $C_{k,j}^f$, the processing sequence of workpieces is denoted as π^f, the overall scheduling scheme is denoted as $\pi = [\pi^1, \pi^2, \dots, \pi^F]$, then the objective function is:

$$C_{max}(\pi) = max\{C_{max}(\pi^f)\} \qquad f = 1,2, \dots, F \tag{1}$$

3. Used methods

3.1 Multi-agent DRL

Multi-agent RL needs to be described by Markov game, because at a same moment, multiple agents respectively take independent actions, after the actions are performed, the reward received by each agent is not only related to itself, but also related to other agents, there're mutual game relationships among agents, and the multi-agent RL system can be described in the following form:

$$(N, S, A, T, R, \gamma) \tag{2}$$

where, N represents the number of multi-agents; S represents the state of the system; A represents the set of actions of the agents, which is consisted of $a_1, a_2, \dots, a_N$; T represents the state transition function, namely $T: S \times A \times S \rightarrow [0,1]$, under state S, after action A is performed, the probability distribution of the next state is obtained; R represents the reward obtained by the agents, that is, $R_i(S, A, S')$ represents the reward obtained by agent i for performing action A and transforming from state S to state S'; γ represents the learning rate. Each agent performs different actions, and all actions constitute systematic joint actions, which make the environment undergo changes and reach a new state, and each agent gets its own reward value, as shown in Fig. 2.

Each agent in the multi-agent system needs to interact with the environment to obtain reward values, thereby improving the strategy and obtaining the optimal strategy in the current environment. The process of the strategy learning is the multi-agent RL.

The matrix game can be expressed as $(N, A_1, A_2, \dots, A_N, R_1, R_2, \dots, R_N)$, where N is the number of agents in the system, A_n is the action set of the n-th agent, and R_n is the reward function of the n-th agent, apparently, the reward obtained by each agent is related to the joint actions of the system, $A_1 \times A_2 \times, \dots, \times A_N$ represents the space of joint actions. The strategy of each agent is the probability distribution of the action space, and the goal of each agent is to get the maximum reward value. Assume π_n represents the strategy of agent n, $(\pi_1, \pi_2, \dots, \pi_N)$ represents the joint strategy of the system, then the value function of agent n can be expressed as $V_n(\pi_1, \pi_2, \dots, \pi_n, \dots, \pi_N)$.

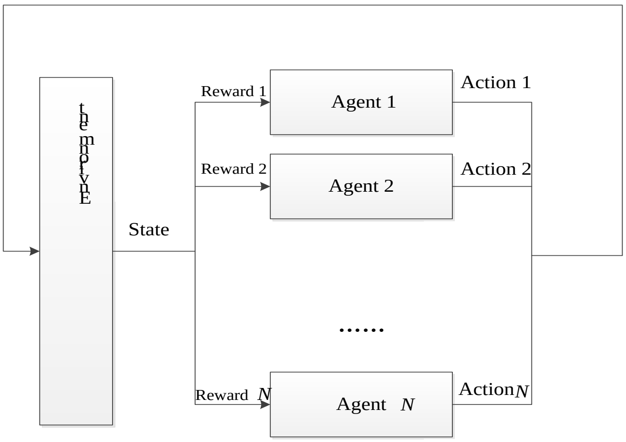


Fig. 2 Multi-agent RL system

If there is $\pi_n \in \prod_n, n = 1,2,\ldots,N$ in the joint strategy $(\pi_1^*, \pi_2^*, \ldots, \pi_N^*)$ of the matrix game, and it satisfies $(\pi_1^*, \pi_2^*, \ldots, \pi_n, \ldots, \pi_N^*) \le (\pi_1^*, \pi_2^*, \ldots, \pi_N^*)$, then it is a NASH equilibrium.

In multi-agent RL algorithm, the NASH equilibrium can be described as:

When the system performs joint actions $[a_1, a_2, \ldots, a_N]$, the expected reward obtained by agent *n* is $R_n[a_1, a_2, \ldots, a_N]$, $\pi_n(a_n)$ represents the probability of agent *n* choosing action a_n, then the NASH equilibrium can be defined as:

$$\sum_{a_1,\ldots,a_N \in A_1,\ldots,A_N} R_n(a_1, a_2, \ldots, a_N)\pi_1^*(a_1)\ldots\pi_n(a_n)\ldots\pi_N^*(a_N) \le \sum_{a_1,\ldots,a_N \in A_1,\ldots,A_N} R_n(a_1, a_2, \ldots, a_N)\pi_1^*(a_1)\ldots\pi_n^*(a_n)\ldots\pi_N^*(a_N)$$

wherein, $\pi_n \in \prod_n, n = 1,2,\ldots,N$.

The Minmax-Q algorithm is an earlier multi-agent method for random decision-making [35]. The algorithm solves the zero-sum game of two agents, and theoretically proves the convergence of the algorithm. However, due to the limitation in actual applications, the algorithm can hardly solve general sum problems. Aiming at this shortcoming of the Minmax-Q algorithm, the NASH-Q algorithm [36] extends to the multi-agent complete information general sum random game problems, and it uses NASH equilibrium to define the value function.

The random game is the random game process of *N* agents, it can be expressed as Eq. 3:

$$(N, S, A_1, A_2, \ldots, A_N, T, R_1, R_2, \ldots, R_N, \gamma) \tag{3}$$

where, *N* represents the number of multi-agents, *S* represents the state of the system, A_n represents the action space of agent *n*, *T* represents the state transition function, namely $T: S \times A \times S \to [0,1]$, under state *S*, after performing action *A*, the probability distribution of the next state could be obtained, R_n represents the reward obtained by agent *n*, γ represents the learning rate, the strategy of agent *n* is defined as $\pi_n: S \to \Omega(A_n)$, which represents the probability distribution from the state to the agent action space, wherein $\Omega(A_n)$ represents all possible sets of the action space probability distribution of agent *n*.

The cumulative discount value function of agent *n* under the current strategy π can be described by Eq. 4 as:

$$V_\pi^n(s) = \sum_{t=0}^{\infty} \gamma^t E_\pi\,[R_t^n | s = s_0, \pi] \tag{4}$$

The action value function can be described as Eq. 5:

$$Q_\pi^n(s, A) = R_n(s, A) + \gamma E_{s' \sim T}[V_\pi^n(s')] \tag{5}$$

The value function described by Eq. 5 is based on the condition of single agent, the action A in the formula is the joint action adopted by each agent, obviously, the action value function of the agents is based on the system state and the joint action space.

Integrating Eq. 5, Eq. 4 can be rewritten as Eq. 6:

$$V_\pi^n(s) = E_{A\sim\pi(s)}[Q_\pi^n(s, A)] \tag{6}$$

It represents that the state value function is obtained through the expectation of the action value function.

Under the state that each agent performs discrete actions, the multi-agent RL is modeled through random games. Under the condition that each agent in the system doesn't know the rewards of other agents, still each agent can observe the previous behaviors of other agents, and respond to the instant rewards generated.

The MiniMax-Q learning algorithm is usually used to solve zero-sum game problems, its core idea is that each agent maximizes the expected reward value in the worst case in the game with the opponent, it solves the Nash equilibrium strategy of the game under specific state s by constructing MiniMax linear programming method, and uses the time-series difference method to iteratively learn the state value function or the action value function. In the zero-sum game of two agents, for a given state s_0, the state value function of the n-th agent can be defined as Eq. 7:

$$V_n^*(s_0) = \max_{\pi_n(s_0,\cdot)} \min_{a_{-n}\in A_{-n}} \sum_{a_n\in A_n} Q_n^*(s_0, a_n, a_{-n})\pi_n(s_0, a_n) \tag{7}$$

where, the value of n is 1 or 2, and a_{-n} represents actions other than a_n.

This paper aims to use multi-agent method to solve the distributed production scheduling problem, therefore, based on the MiniMax-Q learning algorithm, it extends to the general sum game problem of multi-agents, namely the NASH Q-learning algorithm, which uses quadratic programming method to solve the NASH equilibrium point, in the game of each state, it could find out the global optimal point or the saddle point, so that the system can converge to the NASH equilibrium point in the cooperative equilibrium or the adversarial equilibrium.

3.2 Multi-agent MF-DRL algorithm

Mean Field Theory is a method for studying complex multi-agent problems using machine learning and physical field theories [37], it can simplify random models with huge scale or complex structure.

The overall goal of using multiple agents to solve distributed production scheduling problems is to minimize the completion time. Each agent must learn the optimal strategy to cooperate with the overall goal of the system, and the solution algorithm was improved based on the work of Ruiz *et al.* [38]. It's assumed that, π represents the joint strategy of the system, π_i represents the strategy of agent i, v_i represents the value function of agent i, N represents the number of agents; π^* represents the joint strategy of the system under NASH equilibrium, and strategy π^* is consisted of N agent strategies $(\pi_1^*, \pi_2^*, \dots, \pi_N^*)$; the system state is s, then the strategy of agent i is π_i^*, and the strategy of other agents except for agent i in the joint strategy π^* is represented as π_{-i}^*, namely $\pi_{-i}^* \doteq (\pi_1^*, \pi_2^*, \dots, \pi_{i-1}^*, \pi_{i+1}^*, \dots, \pi_N^*)$, and the value function of agent i can be described by Eq. 8:

$$v_i(s; \pi^*) = v_i(s; \pi_i^*, \pi_{-i}^*) \tag{8}$$

According to the feature of NASH equilibrium, Eq. 9 could be obtained as:

$$v_i(s; \pi_i^*, \pi_{-i}^*) \geq v_i(s; \pi_i, \pi_{-i}^*) \tag{9}$$

The multi-agent system is in the optimal state at this moment, in the system, agent i executes tasks according to strategy π_i^*, and other agents execute tasks according to strategy π_{-i}^*, obviously, $i \in \{1,2, \dots, N\}$, that is, any agent in the system has the opportunity to act as an independent agent, and each agent has the opportunity to act as a virtual agent and work with other agents to play the game with independent agents in the system.

The initial state of the system is s, then the value function of the system can be described by Eq. 10:

$$v^*(s) \doteq [v_1(s), v_2(s), \ldots, v_N(s)]_{\pi^*} \tag{10}$$

The Q-value of the current state is initialized according to the definition of the NASH Q-learning algorithm, and is updated continuously through iterations, also, a NASH factor is added during Q-value update to ensure that the Q-value update can satisfy the compression mapping requirements. At the same time, the relationship between the Q-value function and the state value function can be known from Eq. 6, that is, the state value function can be updated synchronously and reach the NASH equilibrium.

In a multi-agent system, it can be known from Eq. 3 that the dimension of the contact action is determined by the number of agents, for any agent i in the system, its action value function $Q_i(s, A)$ can be defined as:

$$Q_i(s, A) = \frac{1}{N-1} \sum_{j \in N-1} Q_i(s, A_i, A_j) \tag{11}$$

This definition makes agent i have a connection with action A_j of other agents, and it simplifies the relationships among multiple agents and gives mathematical descriptions. The complex relationships among multiple agents are simplified as pairwise correspondence between agent i and other agents, then by summing and averaging, the internal interactive relationships between agent i and other agents, and between other agents are retained; the pairwise approximation not only reduces the complexity of the interaction between agents, but also implicitly preserves the global interaction between any pair of agents.

In the workshop production scheduling problems, the sorting of workpieces within production line and the exchange of workpieces between production lines are both discrete actions. Assume: agent i has a total of χ actions, the action space of A_i has a total of χ components, and each component represents a selectable action. A_i is encoded by the One-Hot encoding method, the position code of the selected action is 1, and the position of other components is 0.

In the algorithm, the distance or similarity between features is calculated by One-Hot encoding, the values of discrete actions are extended to the Euclidean space, and then a certain value of the discrete action can correspond to a certain point in the Euclidean space. At the same time, One-Hot encoding is also adopted for discrete features, which makes the calculation of distances between actions more reasonable. After discrete actions are subject to One-Hot encoding, the features of each dimension can be regarded as continuous features, and can be normalized using the normalization method of continuous features, for instance, they can be normalized to [-1,1] or to a mean of 0 and a variance of 1. The characteristics of One-Hot encoding provide convenience for calculating the mean action in the mean field theory, assuming $\overline{A_i}$ represents the mean action, which is equivalent to the multinomial distribution of the actions of agents in the agent neighborhood, then it can be expressed by Eq. 12:

$$\overline{A_i} = \frac{1}{N-1} \sum_j A_j \tag{12}$$

where, A_j represents the neighborhood agent action of agent i, and it can be expressed by Eq. 13:

$$A_j = \overline{A_i} + \Delta A_{i,j} \tag{13}$$

where, $\Delta A_{i,j}$ represents the distance between the neighborhood agent action of agent i and the mean action code.

Then on this basis, the paired Q functions in Eq. 11 are analyzed, the first-order approximation of the action value Q_i of the agent is used to convert the interaction of multiple agents into the interaction between two agents (as the number of neighborhood agents increases, the accuracy increases, it's because the average value of its high-order terms approaches 0), namely the virtual agent that is composed of an agent and its neighborhood agents.

Thus, Eq. 11 can be rewritten as Eq. 14:

$$Q_i(s,A) = Q_i(s,A_i,\overline{A_i}) \tag{14}$$

Therefore, the multi-agent interaction is converted into two-agent interaction, that is, the mean field theory is applied to simplify the pairwise interaction between an agent i and its neighborhood agents to the interaction between the agent i and the mean value of its neighborhood agents.

After the agent system performs action A_i, the system state transits from s to s', and an instant reward R_i is obtained, then the mean field action value function update formula of the system can be described by Eq. 15 as:

$$Q_i^{t+1}(s,A_i,\overline{A_i}) = Q_i^t(s,A_i,\overline{A_i}) + \eta[R_i + \gamma v_i^t(s') - Q_i^t(s,A_i,\overline{A_i})] \tag{15}$$

where, η represents the learning rate, γ represents the discount factor.

In Eq. 15, the mean field value function is used to replace the commonly used maximum value function to iteratively solve the action value function. The reason is that if the maximum value function is adopted, the cooperation of neighborhood agent strategies is required, and the central agent cannot directly change the strategies of neighborhood agents. Besides, if each agent is greedy to obtain actions, then eventually the algorithm will fail to converge due to the dynamic instability of the environment.

The definition of the state value function is extended to the mean field state value function, which can be expresses by Eq. 16:

$$v_i^t(s) = \sum_{A_i} \pi_i^t(A_i|s,\overline{A_i}) E_{\overline{A_i}(A_{-i}\sim\pi_{-i}^t)}[Q_i^t(s,A_i,\overline{A_i})] \tag{16}$$

In the staged game at each moment, $\overline{A_i}$ is obtained via strategy π_j^t of neighborhood agent j at the previous moment, and strategy π_j^t is also described using parameters of the mean action $\overline{A}_j^{t-1}$ of the previous moment, its update process is shown as Eq. 17:

$$\overline{A}_i = \frac{1}{N-1}\sum_j A_j \quad \text{and} \quad A_j \sim \pi_j^t(\cdot\,|s,\overline{A}_j^{t-1}) \tag{17}$$

The strategy π in Eqs. 16 and 17 can also be regarded as the probability distribution of actions taken by an agent. Obviously, the probability that an agent i takes the action A_i at time moment t depends on the current state and the mean field action $\overline{A_i}$, and it can be described as Eq. 18. The agent strategy is a random behavior in which the action obeys a certain probability distribution. The agent influences itself by observing the history behaviors of other agents in the neighborhood. Under a new state, an agent can determine its own best response action based on the history behaviors of other agents. The strategies of other agents in the neighborhood also obey certain probability distribution rules, and the probability distribution can be determined based on prior knowledge and observed values, that is, to determine the strategy. Therefore, by observing the history behaviors of other agents in the learning process, an agent can learn the strategies of other agents and its influence on the system could be obtained.

$$\pi_i^t(A_i|s,\bar{A}_i) = \frac{\exp[-\lambda Q_i^t(s,A_i,\overline{A_i})]}{\sum_{A_i\in A}\exp[-\lambda Q_i^t(s,A_{i'},\overline{A_i})]} \tag{18}$$

Eqs. 17 and 18 are continuously and interactively updated, which can continuously improve the performance of the strategy and obtain the largest cumulative reward value. In addition, an exploratory factor λ is added to Eq. 18 to try different behaviors by sacrificing some short-term benefits, that is, for each current state, with a certain probability, behaviors that have not been tried before in the current state are tried to collect more information so that the agent could reach the optimal strategy in the macroscopic scale. After the agent takes corresponding actions in the current state, it will observe the new state of the system after the joint action is executed, and timely correct the credibility of other agents in the current state.

In a multi-agent system, the learning of any agent should continuously interact with the learning of other agents and continuously modify parameters to achieve the maximum cumulative reward value of the system, and the relationships among agents should be established through certain methods. The actions taken when the system transits from the current state to the next state should be jointly determined by the joint actions of the agents, which indirectly realizes the communication between agents.

To ensure the flow of workpieces between different production lines, the action value function of each agent is fitted by the deep neural network, and the loss function is constructed as Eq. 19:

$$L(\theta_i^q) = [y_i - Q_{\theta_i^q}(s, A_i, \overline{A_i})]^2 \tag{19}$$

where, θ_i^q represents the parameters of the agent, $Q_{\theta_i^q}(s, A_i, \overline{A_i})$ represents the action value function fitted by the deep neural network, y_i represents the target value of the mean field value function and is calculated by Eq. 20:

$$y_i = R_i + \gamma v_{\theta_{-i}^q}(s') \tag{20}$$

where, R_i is the reward of agent i, θ_{-i}^q represents the calculation parameter of the mean field value function.

By taking the derivative of Eq. 19, the gradient direction of the parameter could be obtained as shown in Eq. 21:

$$\nabla_{\theta_i^q} L(\theta_i^q) = [y_i - Q_{\theta_i^q}(s, A_i, \overline{A_i})] \nabla_{\theta_i^q} Q_{\theta_i^q}(s, A_i, \overline{A_i}) \tag{21}$$

The parameters can be updated by solving the gradient descent method.

During application, each agent represents a production line. Within the production lines, for any agent i, the action value function under parameters θ_i^q and θ_{-i}^q are initialized and the corresponding mean action $\overline{A_i}$ of the agent is calculated.

The outer loop for sample collection and the inner loop for training are designed. When the collected samples meet the batch size requirement, it enters the inner loop and realizes more efficient learning through the limited samples collected via the interaction between the agents and the environment, as shown in Fig. 3.

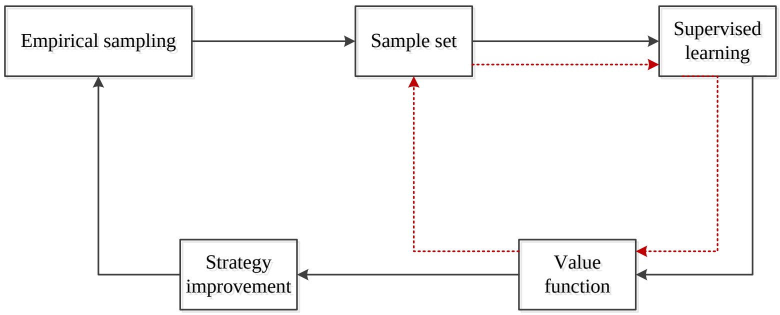


Fig. 3 Algorithm framework

Through the alternation of the two stages: sample collection and storage, and using state action value function to approximate the deep neural network, in the inner loop of the algorithm, the learning algorithm is strengthened through the Q-learning model-free batch processing mode, and it is extended in the multi-agent RL system and iterated by fitting Q. The purpose is to calculate the optimal strategy, and approximate the value function corresponding to the input state and action.

Construct a quintuple as Eq. 22:

$$Tuples = \{(s^t, A^t, \overline{A}^t, R^t, s'^t) | t = 1,2, \dots, |Tuples|\} \tag{22}$$

where, s represents the current state, A represents the action performed by the agent, $\overline{A}$ represents the mean action, R represents the instantaneous reward value, and s'^t represents the new state after performing the action. In the outer loop, the algorithm continuously collects and stores samples according to the element information determined by the quintuple, that is:

$$Set = \{(in^t, out^t) | t = 1,2, \dots, |Tuples| \tag{23}$$

where,

$$in^t = (s^t, A^t, \overline{A}^t) \tag{24}$$

$$out^t = R^t + \gamma \max Q^{counter-1}(s'^t, A', \overline{A}') \tag{25}$$

where, *counter* represents the counter constant.

The information of the quintuple is taken as the input, in each loop, after the value function and the counter are initialized as 0, the approximate value of $Q^{counter}$ is calculated through the training set and the regression algorithm. On this basis, a deep neural network is built as the value function approximator to achieve efficient interaction with the environment, each agent calculates and improves via the defined value function to generate high-quality strategies.

When the discrete action space is very large or it is a continuous action space, the deep neural network is taken as the value function approximator to calculate the value function of the state and action $\tilde{Q}(s, A_i, \overline{A_i})$, at the same time, the Bellman equation is used to calculate $R + \gamma \max Q(s', A', \overline{A}')$, and an error function is introduced to calculate the deviation between the two. The back propagation algorithm is used to calculate the connection weight of the deep neural network, so that the error function value is minimized.

$$\left(\tilde{Q}(s,a) - (r + \gamma\, max_{b \in A(s')} \tilde{Q}(s', b))\right)^2 \tag{26}$$

In a multi-agent system, each agent must go through two stages of individual learning and collaborative improvement. For each agent, the specific state, action, mean action are taken as the inputs of the deep neural network, the estimated corresponding state action value function is taken as the output, then the agent will select the subsequent action according to the instruction of the output value. Under the batch processing mode, in each sampling stage, the RL method initializes the state action value function, in the t sampling steps, if the final state s^{final} is not achieved, then the collected sample is stored into set *Tuples*; if it is the final state and the number of samples meets the requirement of |*Tuples*|, then the value function is updated; otherwise, if the number of samples doesn't meet the requirement of |*Tuples*|, then executes the greedy strategy to choose action, and proceed the sampling.

The outer loop in Fig. 3 describes the interaction between the system and the agent whose learning is reinforced under batch processing mode based on the value function. In the algorithm, the orderly dependency between the action set and state is updated, then, it describes the specific realization process of the outer loop of RL framework under the batch processing mode. By constructing the number of tuples that meets the threshold requirement, after sample collection and storage is completed, the algorithm determines whether to enter the inner loop or not, and then executes the value function update of the batch processing mode.

Samples are used to train the deep neural network value function approximator and calculate the out^t of the current mode, the deviation under the batch learning mode can be expressed by Eq. 27:

$$\sum_{t=1}^{|Tuples|} (Q(s^t, A^t, \overline{A}^t) - out^t)^2 \tag{27}$$

Obviously, the optimization goal of the connection parameters of the deep neural network is:

$$Q^* \leftarrow arg\ \min \sum_{t=1}^{|Tuples|} (Q(s^t, A^t, \overline{A}^t) - out^t)^2 \tag{28}$$

In the process of neural network training, the RMSProp algorithm is adopted, this algorithm can well solve the change range of the parameters after update in the optimization, namely the swing amplitude problem. The RMSProp algorithm uses differential squared weighted averages through the gradients of weights and biases, which is conductive to eliminating the direction with large swing amplitude, and correcting the swing amplitude so that the swing amplitude of each dimension is smaller and the network converges faster. At the same time, the RMSProp algorithm can be built on batch processing training data naturally, and it is easier to be integrated into a batch-processing mode RL algorithm. In this way, the approximator can not only be used for the RL of a single agent, through the RMSProp neural network training algorithm, it also realizes high-quality neural network approximation value function, and it is the key link for the algorithm to be extended to the multi-agent system.

4. Results and discussion of case studies

4.1 Experiment setup

In this study, the experiment constructed 9 states and 10 actions, the features of the states were described by the processing status of the processing machines and the completion status of the workpieces, such as the workload of the processing machines, the estimated earliest possible completion time, and the estimated maximum completion time, etc.

This paper took the minimization of the maximum completion time as the goal to research the workshop scheduling problem, and the global reward was expressed as the sum of local rewards, in this way, the learning system was regarded as a multi-agent learning system with independent rewards, further, in the algorithm, attentions were first paid to the local rewards of each agent obtained according to local scheduling strategies, and then to the global reward. To intuitively understand the characteristics of the distributed scheduling problem, first, the sorting inside each production line should be completed, if the equilibrium condition is not reached, then execute the step of workpiece reallocation between production lines and other steps.

In workshop scheduling, the processing machines were not allowed to perform two operations within a discrete time step. Once a machine starts to perform a certain procedure on the workpiece, it will be in the working state until the end of the procedure. Therefore, for an agent i, its value function was calculated according to the machine state at the decision time point t and the next decision time point $t + \Delta t$.

In the sample collection stage of the agent, the batch size was set to be 100, which means to enter the inner loop of the algorithm when the sample size reaches 100. About the supervised learning part, in the experiment, for the relevant parameters of the RMSProp algorithm, the multifactor varianceanalysis program was employed to analyze the obtained experimental results and determine the relationships among parameters and generate the best parameter combination.

The parameters were respectively set as

$$dr = (0.75,0.85,0.95,0.99),\ lr = (10^{-4}, 10^{-5}, 10^{-6}), sc = (10^{-4}, 10^{-5}, 10^{-6}),$$
$$epoch = (600,800,1000,1200),\ iw = ([-0.06,0.06], [-0.08,0.08], [-0.10,0.10]),$$

and the 5 groups of parameters had a total of $4 \times 3 \times 3 \times 4 \times 3$=432 combinations. After analyzing the experimental results using the variance analysis method, the p-values of the 5 groups of parameters were all lower than the 0.05 confidence interval, indicating that the algorithm was sensitive to these parameters, and finally it's determined that the decay rate of RMSProp algorithm was 0.95, the learning rate lr was 10^{-5}, the small constant sc was 10^{-6}, the epoch of the neural network algorithm was 1000, and the initial weight iw was a random number that is uniformly distributed between [-0.08, 0.08].

The influence of parameters on algorithm performance is shown in Fig. 4.

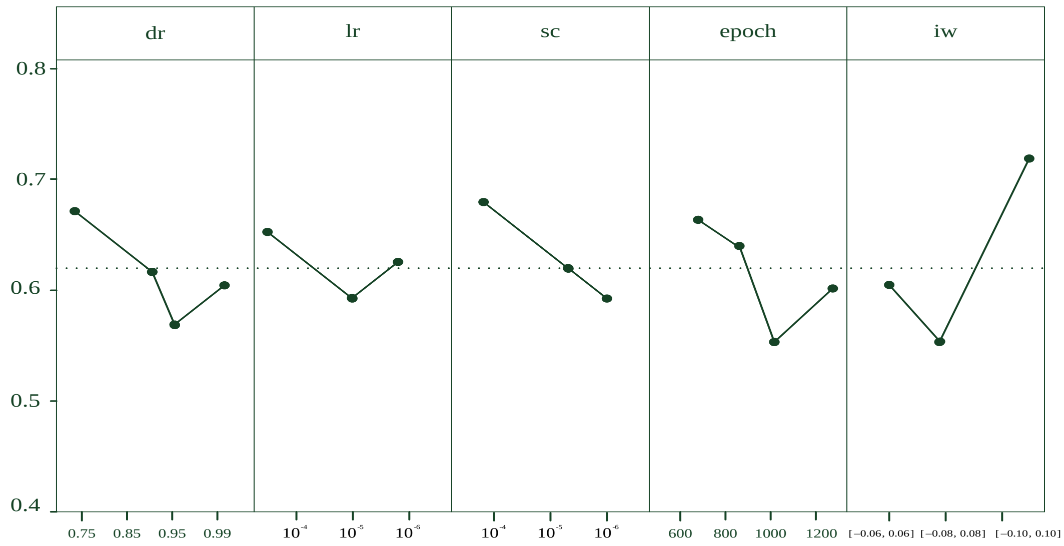


Fig. 4 Influence of parameters on algorithm performance

4.2 Comparison and analysis of results

Based on the benchmark test set Taillard, the performance of the multi-agent RL algorithm was tested, and the results were compared with the results of the iterative greedy algorithm and the benchmark results. In addition, Python was adopted to implement offline training of deep RL and the iterative greedy algorithm was implemented in a Matlab environment. By calculating the average relative percentage deviation (ARPD) index of each test question, the results were compared, as shown in Eq. 29:

$$ARPD = \frac{1}{NR}\sum_{nr=1}^{NR}\frac{C_{max} - C^*_{max}}{C^*_{max}} \times 100 \tag{29}$$

where, NR represents the number of test runs, C_{max} represents the minimum completion time obtained in the nr-th experiment, and C^*_{max} represents the known optimal completion time. The smaller the value of $ARPD$, the better the performance of the algorithm. In the experiment, the number of agents was respectively set as 2, 3, 4, 5, 6, 7, 8, 9, 10 for testing, and the time level was respectively set as 20, 40, and 60. The ARPD comparison results are shown in Tables 1-3.

The calculation results of iterative greedy algorithm (IG), multi-agent RL algorithm (MARL), IG1S, and IG2S [39] were compared. In the small-scale benchmark test, a total of 60 benchmark questions ($20 \times 5, 20 \times 10, 20 \times 20, 50 \times 5, 50 \times 10, 50 \times 20$) in the test set Taillard were selected, and three time levels T = 20, T = 40, and T = 60 were set in the experiment. When T = 60, the average value of ARPD of the MARL algorithm was the smallest among the four algorithms, followed by IG2S, the average ARPD of the IG algorithm designed in this paper was between those of the IG2S and the IG1S, the average ARPD values of the four algorithms were close, and the average ARPD of the IG2S algorithm was 12 % higher than that of the MARL algorithm. When T = 40, the average ARPD values of the four algorithms declined to varying degrees, indicating that within a longer computation time, the algorithms could obtain better experimental results; the average ARPD values of the four algorithms respectively decreased by 5 %, 9 %, 6 %, and 7 %; the IG2S algorithm had the largest average ARPD decrement, followed by the MARL algorithm, whose overall average ARPD still kept the lowest. When T = 60, compared with the results then T = 40, the average ARPD of the algorithms showed no obvious increase, in actual industrial applications, considering the real-time response requirements of the algorithm for the production environment, choosing T = 40 for the algorithm is more appropriate.

Table 1 ARPD comparison (T = 20)

Number of production lines	IG1S	IG2S	IG	MARL
2	0.68	0.52	0.64	0.53
3	0.69	0.60	0.66	0.55
4	0.68	0.63	0.67	0.56
5	0.70	0.64	0.69	0.56
6	0.72	0.66	0.68	0.58
7	0.72	0.68	0.72	0.58
8	0.93	0.79	0.89	0.59
9	0.99	0.89	0.89	0.60
10	1.21	1.05	1.01	0.64
Mean	0.81	0.72	0.76	0.58

Table 2 ARPD comparison (T = 40)

Number of production lines	IG1S	IG2S	IG	MARL
2	0.62	0.48	0.54	0.46
3	0.63	0.49	0.54	0.48
4	0.65	0.52	0.59	0.50
5	0.65	0.52	0.69	0.49
6	0.68	0.66	0.60	0.51
7	0.73	0.68	0.68	0.53
8	0.90	0.73	0.79	0.54
9	0.92	0.83	0.86	0.54
10	1.19	0.92	0.98	0.60
Mean	0.77	0.65	0.70	0.52

Table 3 ARPD comparison (T = 60)

Number of production lines	IG1S	IG2S	IG	MARL
2	0.60	0.46	0.54	0.45
3	0.62	0.46	0.53	0.46
4	0.65	0.50	0.58	0.50
5	0.63	0.50	0.66	0.49
6	0.66	0.65	0.70	0.50
7	0.71	0.66	0.67	0.53
8	0.90	0.73	0.73	0.53
9	0.91	0.81	0.81	0.54
10	1.17	0.91	0.95	0.60
Mean	0.76	0.63	0.69	0.51

The comparison of average ARPD values of the four algorithms at different time levels in the small-scale test is shown in Fig. 5.

In the experiment of large-scale calculation examples, five types of scales ($100 \times 5, 100 \times 10, 100 \times 20, 200 \times 20, 500 \times 20$) and a total of 50 calculation examples were chosen for the tests. Similarly, three time levels T = 20, T = 40, and T = 60 were set in the experiment, and the results of the ARPD values of the four algorithms are shown in Tables 4-6.

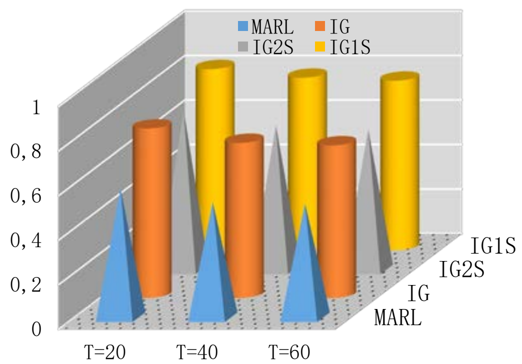


Fig. 5 Comparison of average ARPD of small-scale calculation examples

Table 4 ARPD comparison (T = 20)

Number of production lines	IG1S	IG2S	IG	MARL
2	1.42	1.02	1.37	0.83
3	1.76	1.32	1.66	0.95
4	1.78	1.53	1.79	1.13
5	1.90	1.74	1.94	1.35
6	2.54	2.34	2.44	1.67
7	2.98	2.58	2.68	1.68
8	3.55	2.87	3.34	1.89
9	3.89	3.01	3.90	1.90
10	4.34	3.05	4.39	2.21
Mean	2.68	2.16	2.61	1.51

Table 5 ARPD comparison (T = 40)

Number of production lines	IG1S	IG2S	IG	MARL
2	1.40	1.00	1.35	0.73
3	1.74	1.30	1.64	0.84
4	1.72	1.51	1.77	1.12
5	1.90	1.70	1.94	1.34
6	2.51	2.33	2.43	1.62
7	2.96	2.57	2.67	1.63
8	3.59	2.84	3.33	1.80
9	3.89	3.00	3.90	1.90
10	4.32	3.02	4.38	2.19
Mean	2.67	2.14	2.60	1.46

Table 6 ARPD comparison (T = 60)

Number of production lines	IG1S	IG2S	IG	MARL
2	1.40	1.00	1.34	0.69
3	1.73	1.27	1.64	0.80
4	1.72	1.49	1.77	1.07
5	1.90	1.70	1.92	1.31
6	2.50	2.31	2.43	1.62
7	2.96	2.53	2.67	1.62
8	3.58	2.84	3.31	1.80
9	3.89	3.00	3.90	1.88
10	4.31	3.02	4.36	2.16
Mean	2.67	2.13	2.59	1.44

In the three large-scale experiments, the workpieces and processing machines increased significantly, and the ARPD values of the algorithms had been improved greatly, indicating that when dealing with large-scale problems, the algorithms' problem-solving abilities had reduced. Taking T = 40 as an example, by comparing the ARPD values of the algorithms in small-scale and large-scale scenarios, the MARL algorithm had the smallest increment, followed by the IG2S algorithm, which indicated that when dealing with large-scale problems, the multi-agent RL algorithm designed in this paper also showed obvious superiority.

In the experiment of large-scale calculation examples, the comparison of average ARPD values of the four algorithms at different time levels is shown in Fig. 6, obviously, the ARPD of the MARL algorithm was the smallest.

In order to further test the performance of the algorithm to solve a larger scale and keep other conditions unchanged, when the number of production lines increases to 15 and 20, it can be found that the performance of the proposed algorithm decreases significantly and loses its leading edge compared with other algorithms. This shows that the algorithm has great limitations in solving large-scale problems. How to further improve the algorithm and improve the performance of the algorithm in solving large-scale problems is a problem we are currently exploring.

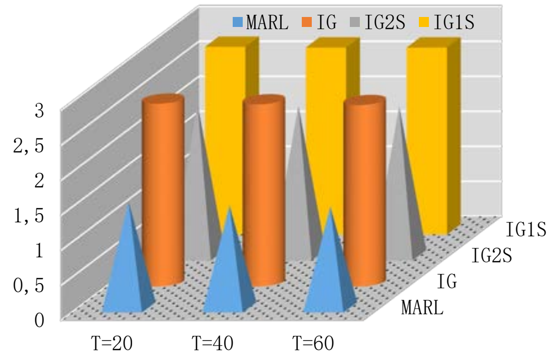


Fig. 6 Comparison of mean ARPD in large-scale computation examples

This paper designed a multi-agent RL method to solve DPFSPs. Based on the NASH equilibrium theory and the NASH Q-learning method, a multi-agent MF-DRL algorithm had been proposed in the study, and global perspective algorithm elements such as joint state and joint actions were constructed. The experimental results showed that, the proposed multi-agent RL method was effective in solving DPFSPs, and it outperformed other algorithms in case of large-scale problems. The RL method proposed in this paper was mainly based on the value function method, it hadn't involved the strategy-based RL method, and now the effect of the strategy-based RL method on the solution of distributed production scheduling hasn't been explored yet, studies of this aspect will be carried out in the subsequent research.

5. Conclusion

This paper designed a multi-agent RL method to solve DPFSPs. Based on the NASH equilibrium theory and the NASH Q-learning method, a multi-agent MF-DRL algorithm had been proposed in the study, and global perspective algorithm elements such as joint state and joint actions were constructed. The experimental results showed that, the proposed multi-agent RL method was effective in solving DPFSPs, and it outperformed other algorithms in case of large-scale problems. The RL method proposed in this paper was mainly based on the value function method, it hadn't involved the strategy-based RL method, and now the effect of the strategy-based RL method on the solution of distributed production scheduling hasn't been explored yet, studies of this aspect will be carried out in the subsequent research.

Acknowledgment

Project supported by the Key Soft Science Project of "Science and Technology Innovation Action Plan of Shanghai Science and Technology Commission, China (No. 20692104300). National Natural Science Foundation, China (No. 71840003), and the Technology Development Project of University of Shanghai for Science and Technology, China (No. 2018KJFZ043).

References

[1] Behnamian, J., Fatemi Ghomi, S.M.T. (2016). A survey of multi-factory scheduling, *Journal of Intelligent Manufacturing*, Vol. 27, No. 1, 231-249, doi: 10.1007/s10845-014-0890-y.
[2] Somashekhara, S.C.H., Setty, A.K.Y., Sridharmurthy, S.M., Adiga, P., Mahabaleshwar, U.S., Lorenzini, G. (2019). Makespan reduction using dynamic job sequencing combined with buffer optimization applying genetic algorithm in a manufacturing system, *Mathematical Modelling of Engineering Problems*, Vol. 6, No. 1, 29-37, doi: 10.18280/mmep.060104.
[3] Zhang, Y.Q., Zhang, H. (2020). Dynamic scheduling of blocking flow-shop based on multi-population ACO algorithm, *International Journal of Simulation Modelling*, Vol. 19, No. 3, 529-539, doi: 10.2507/IJSIMM19-3-CO15.
[4] Wang, L., Deng, J., Wang, S.-Y. (2016). Survey on optimization algorithms for distributed shop scheduling, *Control and Decision*, Vol. 31, No. 1, 1-11.
[5] Fernandez-Viagas, V., Framinan, J.M. (2015). A bounded-search iterated greedy algorithm for the distributed permutation flowshop scheduling problem, *International Journal of Production Research*, Vol. 53, No. 4, 1111-1123, doi: 10.1080/00207543.2014.948578.

[6] Gao, J., Chen, R. (2011). An NEH-based heuristic algorithm for distributed permutation flowshop scheduling problems, *Scientific Research and Essays*, Vol. 6, No. 14, 3094-3100.

[7] Gao, J., Chen, R., Deng, W. (2013). An efficient tabu search algorithm for the distributed permutation flowshop scheduling problem, *International Journal of Production Research*, Vol. 51, No. 3, 641-651, doi: 10.1080/00207543.2011.644819.

[8] Naderi, B., Ruiz, R. (2014). A scatter search algorithm for the distributed permutation flowshop scheduling problem, *European Journal of Operational Research*, Vol. 239, No. 2, 323-334, doi: 10.1016/j.ejor.2014.05.024.

[9] Lin, S.-W., Ying, K.-C., Huang, C.-Y. (2013). Minimising makespan in distributed permutation flowshops using a modified iterated greedy algorithm, *International Journal of Production Research*, Vol. 51, No. 16, 5029-5038, doi: 10.1080/00207543.2013.790571.

[10] Yang, S.L., Xu, Z.G., Li, G.Z., Wang, J.Y. (2020). Assembly transport optimization for a reconfigurable flow shop based on a discrete event simulation, *Advances in Production Engineering & Management,* Vol. 15, No. 1, 69-80, doi: 10.14743/apem2020.1.350.

[11] Wang, S.-Y., Wang, L., Liu, M., Xu, Y. (2013). An effective estimation of distribution algorithm for solving the distributed permutation flow-shop scheduling problem, *International Journal of Production Economics*, Vol. 145, No. 1, 387-396, doi: 10.1016/j.ijpe.2013.05.004.

[12] Rifai, A.P., Nguyen, H.-T., Dawal, S.Z.M. (2016). Multi-objective adaptive large neighborhood search for distributed reentrant permutation flow shop scheduling, *Applied Soft Computing*, Vol. 40, 42-57, doi: 10.1016/j.asoc.2015.11.034.

[13] Wang, K., Huang, Y., Qin, H. (2016). A fuzzy logic-based hybrid estimation of distribution algorithm for distributed permutation flowshop scheduling problems under machine breakdown, *Journal of the Operational Research Society*, Vol. 67, No. 1, 68-82, doi: 10.1057/jors.2015.50.

[14] Komaki, M., Malakooti, B. (2017). General variable neighborhood search algorithm to minimize makespan of the distributed no-wait flow shop scheduling problem, *Production Engineering,* Vol. 11, No. 3, 315-329, doi: 10.1007/s11740-017-0716-9.

[15] Chen, W., Hao, Y.F. (2018). Genetic algorithm-based design and simulation of manufacturing flow shop scheduling, *International Journal of Simulation Modelling,* Vol. 17, No. 4, 702-711, doi: 10.2507/IJSIMM17(4)CO17.

[16] Lebbar, G., El Abbassi, I., Jabri, A., El Barkany, A., Darcherif, M. (2018). Multi-criteria blocking flow shop scheduling problems: Formulation and performance analysis, *Advances in Production Engineering & Management,* Vol. 13, No. 2, 136-146, doi: 10.14743/apem2018.2.279.

[17] Rathinam, B., Govindan, K., Neelakandan, B.,Raghavan, S.S. (2015). Rule based heuristic approach for minimizing total flow time in permutation flow shop scheduling, *Tehnički Vjesnik – Technical Gazette,* Vol. 22, No. 1, 25-32, doi: 10.17559/TV-20130704132725.

[18] Babu, K.S., Vemuru, S. (2019). Spectrum signals handoff in LTE cognitive radio networks using reinforcement learning, *Traitement du Signal*, Vol. 36, No. 1, 119-125, doi: 10.18280/ts.360115.

[19] Foerster, J., Nardelli, N., Farquhar, G., Afouras, T., Torr, P.H.S., Kohli, P., Whiteson, S. (2017). Stabilising experience replay for deep multi-agent reinforcement learning, In: *Proceedings of 34th International Conference on Machine Learning,* Sydney, Australia, 1146-1155.

[20] Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., Vicente, R. (2017). Multiagent cooperation and competition with deep reinforcement learning, *PloS ONE*, Vol. 12, No. 4, Article No. e0172395, doi: 10.1371/journal.pone.0172395.

[21] Xu, X., Jia, Y., Xu, Y., Xu, Z., Chai, S., Lai, C.S. (2020). A multi-agent reinforcement learning-based data-driven method for home energy management, *IEEE Transactions on Smart Grid*, Vol. 11, No. 4, 3201-3211, doi: 10.1109/TSG.2020.2971427.

[22] Zerguine, N., Mostefai, M., Aliouat, Z., Slimani, Y. (2020). Intelligent CW selection mechanism based on Q-learning (MISQ), *Ingénierie des Systèmes d'Information*, Vol. 25, No. 6, 803-811, doi: 10.18280/isi.250610.

[23] Sukhbaatar, S., Szlam, A., Fergus, R. (2016). Learning multiagent communication with backpropagation, In: *Proceedings of 30th Conference on Neural Information Processing Systems (NIPS 2016),* Barcelona, Spain, 2252-2260.

[24] Foerster, J.N., Assael, Y.M., de Freitas, N., Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning, In: *Proceedings of 30th Conference on Neural Information Processing Systems (NIPS 2016),* Barcelona, Spain, 2137-2145.

[25] He, H., Boyd-Graber, J., Kwok, K., Daumé III, H. (2016). Opponent modeling in deep reinforcement learning, In: *Proceedings of 33rd International Conference on Machine Learning,* New York, USA, 1804-1813.

[26] Palmer, G., Tuyls, K., Bloembergen, D., Savani, R. (2018). Lenient multi-agent deep reinforcement learning, In: *Proceedings of 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018,* Stockholm, Sweden, 443-451.

[27] Omidshafiei, S., Pazis, J., Amato, C., How, J.P., Vian, J. (2017). Deep decentralized multi-task multi-agent reinforcement learning under partial observability, In: *Proceedings of the 34th International Conference on Machine Learning (ICML 2017),* Sydney, Australia, 2681-2690.

[28] Sunehag, P., Lever, G., Gruslys, A., Czarnecki, W.M., Zambaldi, V., Jaderberg, M., Lanctot, M., Sonnerat, N., Leibo, J.Z., Tuyls, K, Graepel, T. (2018). Value-decomposition networks for cooperative multi-agent learning based on team reward, In: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018),* Stockholm, Sweden, 2085-2087.

[29] Li, C., Tang, J., Ma, T., Yang, X., Luo, Y. (2020). Load balance based workflow job scheduling algorithm in distributed cloud, *Journal of Network and Computer Applications*, Vol. 152, Article No. 102518, doi: 10.1016/j.jnca.2019.102518.
[30] Neary, C., Xu, Z., Wu, B., Topcu, U. (2021). Reward machines for cooperative multi-agent reinforcement learning, In: *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '21,* Virtual Event, United Kingdom, 934-942.
[31] Li, R., Han, Y., Ma, T., Liu, H. (2020). Nash-Q learning-based collaborative dispatch strategy for interconnected power systems, *Global Energy Interconnection*, Vol. 3, No. 3, 227-236, doi: 10.1016/j.gloei.2020.07.004.
[32] Li, J., Xiao, Z. (2020). H_∞ control for discrete-time multi-player systems via off-policy Q-learning, *IEEE Access*, Vol. 8, 28831-28846, doi: 10.1109/ACCESS.2020.2970760.
[33] Shah, D., Somani, V., Xie, Q., Xu, Z. (2020). On reinforcement learning for turn-based zero-sum Markov games, In: *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference, FODS '20,* Virtual Event, USA, 139-148, doi: 10.1145/3412815.3416888.
[34] Feng, Q., Shang, W.-P., Jiao, C.-W., Li, W.-J. (2020). Two-agent scheduling on a bounded parallel-batching machine with makespan and maximum lateness objectives, *Journal of the Operations Research Society of China*, Vol. 8, No. 1, 189-196, doi: 10.1007/s40305-019-00258-9.
[35] Littman, M.L. (1994). Markov games as a framework for multi-agent reinforcement learning, In*: Proceedings of the Eleventh International Conference on International Conference on Machine Learning, ICML'94,* New Brunswick USA, 157-163, doi: 10.1016/B978-1-55860-335-6.50027-1.
[36] Hu, J., Wellman, M.P. (1998). Multiagent reinforcement learning: Theoretical framework and an algorithm, In: *Proceedings of the Fifteenth International Conference on Machine Learning, ICML '98,* Madison, USA, 242-250.
[37] Xiao, L., Bahri, Y., Sohl-Dickstein, J., Schoenholz, S., Pennington, J. (2018). Dynamical isometry and a mean field theory of cnns: How to train 10,000-layer vanilla convolutional neural networks, In: *Proceedings of the 35th International Conference on Machine Learning,* Stockholm, Sweden, 5393-5402.
[38] Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., Wang, J. (2018). Mean field multi-agent reinforcement learning, In: *Proceedings of the 35th International Conference on Machine Learning,* Stockholm, Sweden, 5571-5580.
[39] Ruiz, R., Pan, Q.-K., Naderi, B. (2019). Iterated Greedy methods for the distributed permutation flowshop scheduling problem, *Omega*, Vol. 83, 213-222, doi: 10.1016/j.omega.2018.03.004.

Appendix

The abbreviations used in the article:

RL	Reinforcement Learning
MF	Mean Field
DFSP	Distributed Flow-Shop Scheduling Problem
DPFSP	Distributed Permutation Flow-shop Scheduling Problem
DRL	Deep Reinforcement Learning
MF-DRL	Mean Field Deep Reinforcement Learning
ARPD	Average Relative Percentage Deviation
MARL	Multi agent Reinforcement Learning
RMSProp	Root Mean Square Prop
IG	Iterative Greedy
IG1S	The single stage Iterative Greedy
IG2S	The two stage Iterative Greedy

APEM journal

Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 285–296
https://doi.org/10.14743/apem2021.3.400

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

Using the gradient boosting decision tree (GBDT) algorithm for a train delay prediction model considering the delay propagation feature

Zhang, Y.D.[a,*], Liao, L.[a], Yu, Q.[a], Ma, W.G.[a], Li, K.H.[b]

[a]School of Information Science and Technology, Southwest Jiaotong University, Chengdu, P.R. China
[b]School of management, Xihua University, Chengdu, P.R. China

A B S T R A C T

Accurate prediction of train delay is an important basis for the intelligent adjustment of train operation plans. This paper proposes a train delay prediction model that considers the delay propagation feature. The model consists of two parts. The first part is the extraction of delay propagation feature. The best delay classification scheme is determined through the clustering method of delay types for historical data based on the density-based spatial clustering of applications with noise algorithm (DBSCAN), and combining the best delay classification scheme and the k-nearest neighbor (KNN) algorithm to design the classification method of delay type for online data. The delay propagation factor is used to quantify the delay propagation relationship, and on this basis, the horizontal and vertical delay propagation feature are constructed. The second part is the delay prediction, which takes the train operation status feature and delay propagation feature as input feature, and use the gradient boosting decision tree (GBDT) algorithm to complete the prediction. The model was tested and simulated using the actual train operation data, and compared with random forest (RF), support vector regression (SVR) and multilayer perceptron (MLP). The results show that considering the delay propagation feature in the train delay prediction model can further improve the accuracy of train delay prediction. The delay prediction model proposed in this paper can provide a theoretical basis for the intelligentization of railway dispatching, enabling dispatchers to control delays more reasonably, and improve the quality of railway transportation services.

A R T I C L E   I N F O

Keywords:
Train delay prediction;
Actual train operation data;
Delay type identification;
Delay propagation feature extraction;
Density-based spatial clustering of applications with noise (DBSCAN);
k-nearest neighbor (KNN);
Gradient boosting decision tree (GBDT);
Random forest (RF);
Support vector regression (SVR);
Multilayer perceptron (MLP)

Corresponding author:
ydzhang@swjtu.edu.cn
(Zhang, Y.D.)

Article history:
Received 24 July 2021
Revised 25 October 2021
Accepted 28 October 2021

1. Introduction

With the development of railway network and the growth of passenger travel demand, the utilization rate of railway lines is getting higher and higher. Under the premise of ensuring the safety of train operation, ensuring the punctuality is the key of railway transportation to improve the quality of service. Once the train is delayed, the dispatchers must use dispatch adjustment methods reasonably and scientifically [1]. The accurate prediction of train delays can assist dispatchers to make scientific decisions, and even realize the intelligent dynamic adjustment of train operation plans.

The traditional train delay prediction mainly relies on the work experience and operating skills of dispatchers. Due to the uncertainty of train delays, this method is difficult to reasonably predict the delay before it occurs. On the other hand, the primary delay caused by the interference of external factors will produce a domino effect-like delay propagation effect on the line, which will lead to the secondary delay. However, it is very limited to rely on the experience of dispatchers to predict the secondary delay. With the development of railway informatization, on the basis that the actual operation data of trains can be fully collected and fully processed, the application of big data and machine learning to train delay prediction has important reference value for the development of intelligent dispatching and command work [2]. The machine learning method is based on the actual train operation data and do not require relevant details within the system. The actual data can reflect the relevant factors and their interactions of delays. This method is conducive to revealing the occurrence and propagation of train delays.

This paper proposes a train delay prediction model that considers the delay propagation feature. When the model uses machine learning methods for delay prediction, the delay propagation feature is added to improve the prediction accuracy. The main contributions of this paper include: (1) Design the clustering method of delay types for historical data based on density-based spatial clustering of applications with noise (DBSCAN) and the classification method of delay types for online data based on k-nearest neighbor (KNN); (2) According to the determined delay type, the delay propagation factor is used to quantify the delay propagation relationship and construct as the horizontal and vertical delay propagation feature; (3) Construct a gradient boosting decision tree (GBDT) model to complete the prediction of train delays according to the train operation status feature and the delay propagation feature.

This paper is organized as follows. Section 2 summarizes the current research on train delay prediction. Section 3 describes the problems to be solved. Section 4 describes the overall structure of the train delay prediction model proposed in this paper and describes the design principles of each part in detail. In Section 5, an example is analyzed based on the actual data of train operation. Section 6 summarizes the work of this paper.

2. Related work

The input feature set of the machine learning model will affect the performance of the model. So the selected input feature should have the greatest impact on the output results.

It is a routine consideration to use information related to train characteristics and train status as the input feature set of the prediction model, because these are factors that directly affect train delays. Oneto *et al.* [3] took the section running time, working day/non-working day feature as the input feature set. Wang and Zhang [4] identified the number of delayed trains at each station, the total value of delays of each station and the total value of each train's delays as factors affecting train delays. Shi *et al.* [5] considered the current station and train codes when establishing the feature set. The related historical value of train operation is also related to the delay prediction. Nair *et al.* [6] considered the historical delay value and historical running time of the train at the station as the input feature of model. But the train runs according to the predesigned train operation diagram, therefore, the planned values related to the operation diagram such as the planned running time [7], the planned stopping time of trains and the planned running interval between trains [8, 9] should also be taken into consideration.

To improve the quality of prediction, some studies have begun to consider other factors besides the feature of train operation status. Tang *et al.* [10] used the primary delay time, the number of affected trains and the delay causes as independent variables. Zhang *et al.* [11] based on the secondary delay data, considered the impact of the preceding train on the current train and constructed a model input feature set. Hu *et al.* [12] used a hierarchical clustering algorithm to analyze the delayed trains and based on the results of 4 types of delayed train sequences made subsequent delay time predictions. Zeng *et al.* [13] used cause analysis to infer the delay propagation chain, integrated the train event information with the primary delay and secondary delay information contained in prediction model.

The above literature review shows that when establishing the input feature of prediction model, most studies consider the train operating status information, and also the factors related to the data characteristics and the predicted output. But few studies consider the delay propagation relationship in the delay prediction. In terms of prediction models, using machine learning models to predict delays has a better fitting effect than traditional statistical models [14, 15]. The machine learning model can realize the prediction of delay more accurately and quickly, and at the same time, the output can be stabilized in the case of a large amount of data. Decision tree model [16], random forest model [6, 7, 9, 13], neural network model [8, 17] and SVR model [10] are now widely used in train delay prediction research.

Therefore, this article will establish a GBDT-PF model considers the delay propagation between trains and the delay propagation of the train itself. The model can identify the delay types and obtain the delay propagation relationship. The delay classification scheme can enable dispatchers to understand the law of the occurrence of primary delays, and the delay propagation relationship has a direct impact on the subsequent train delays, so the prediction model that considers the delay propagation relationship can obtain higher accuracy.

3. Problem statement

The train operation process in the railway network can be expressed as a collection of a series of events and processes [18]. The dependency between events and processes can be represented by timed event graphs. Fig. 1 is the use of time event graph to show the operation status of each train. According to the time of day, each train is arranged vertically $(1,2,3,\cdots,i)$ and each station is arranged horizontally $(1,2,3,\cdots,s)$.

In Fig. 1, the node $t_{i,s}$ represents the arrival or departure event of train i at station s, the weight of the node $D_{i,s}$ represents the delay value of train i at station s, the directed arc $arc(t_{i,s}, t_{i_{+1},s})$ represents the operation process that train i and train $i+1$ run to station s. the weight of the directed arc $w(t_{i,s}, t_{i+1,s})$ represents the running interval between train i and train $i+1$ at station s. When $t_{i,s}$ represents the departure event, the directed arc $arc(t_{i,s}, t_{i,s_{+1}})$ represents the running process of train i from station s to station $s+1$, the weight of the directed arc $w(t_{i,s}, t_{i,s_{+1}})$ represents the running time of the corresponding train i from station s to the station $s+1$. When $t_{i,s}$ represents the arrival event, the directed arc $arc(t_{i,s}, t_{i,s_{+1}})$ represents the stopping process of train i at station s, and the weight of the directed arc $w(t_{i,s}, t_{i,s_{+1}})$ represents the stop time of train i at station s.

According to Fig. 1, when $D_{i,s}$ is primary delay, the horizontal propagation of the delay occurs through the directed arc $arc(t_{i,s}, t_{i+1,s})$, and it affects the next train. Vertical propagation of the delay occurs through the directed arc $arc(t_{i,s}, t_{i,s+1})$, and it affects the train itself.

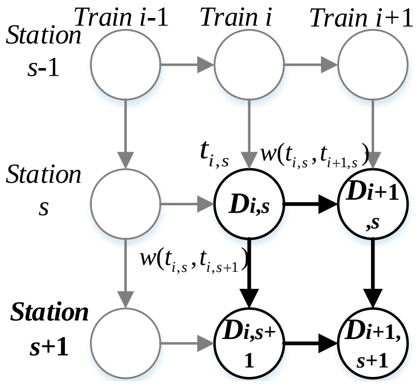


Fig. 1 Time event graph of train operation status (when $t_{i,s}$ represents a departure event)

The delay value of the train should be related to the historical value and has nothing to do with whether the train will be delayed in the future. For the node $t_{i,s}$, there are only two nodes directly related to it, the node $t_{i\text{-}1,s}$ in the horizontal direction and the node $t_{i,s\text{-}1}$ in the vertical direction. These two nodes also represent the delay propagation between trains and the delay propagation of the train itself. Therefore, given the train i and the station s, considering the train operating state factors Z_i, the delay propagation factor of the train itself $P_{i,s-1}$ and the delay propagation factor between the trains $P_{i-1,s}$, a nonlinear function is learned to complete the prediction of $\widehat{D}_{i,s}$:

$$\widehat{D}_{i,s} = f\left(Z_i, P_{i,s-1}, P_{i-1,s}\right) \tag{1}$$

Among them, $\widehat{D}_{i,s}$ is the delay time of the arrival or departure of train i at station s. Z_i is a feature set related to $D_{i,s}$ based on historical data, $P_{i,s-1}$is a feature set related to the vertical propagation of delays, $P_{i-1,s}$ is a feature set related to the horizontal propagation of delays, and f is the establishment machine learning model.

4. Methodology

Fig. 2 shows the structure of the GBDT-PF model. The feature extraction part is to complete the construction of the input feature set. Then the GBDT model is trained to realize the prediction of $\widehat{D}_{i,s}$. For the value $\widehat{D}_{i,s}$, it is necessary to confirm the type of the two nodes $t_{i,s-1}$ and $t_{i-1,s}$, but since the original data set has no related records for the type of delay, before constructing the feature set, the first task is to classify the delay types in the historical data, then construct a data set that can be used for delay type identification, so as to carry out the subsequent delay prediction.

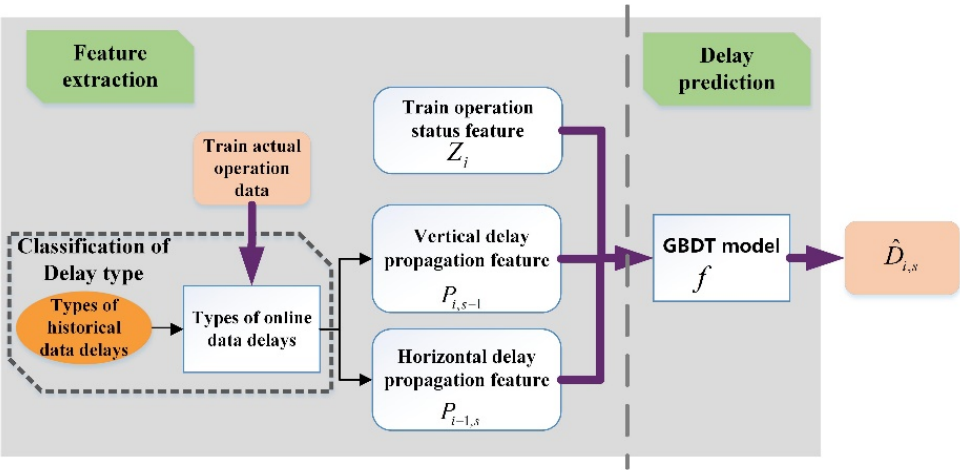


Fig. 2 The GBDT-PF model

4.1 Cluster analysis of delay types for historical data based on DBSCAN

Using DBSCAN algorithm to complete the cluster analysis of delay types for historical data is shown in Fig. 3. For $D_{i,s}$, according to the characteristics of the primary delay and the secondary delay, the input feature set of the cluster analysis is determined as shown in Table 1.

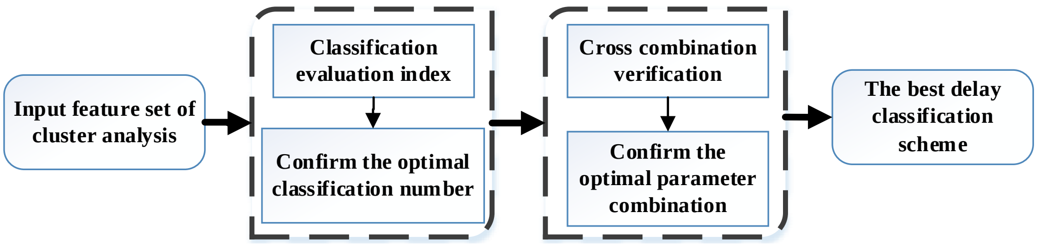


Fig. 3 The process of cluster analysis of delay types for historical data based on DBSCAN

Table 1 The input feature set of the cluster analysis

Feature	Symbol	Meaning
Input feature set of cluster analysis	$Tdelay_{i,s}$	The delay value of train i at station s, a positive value means delay, a negative value means early, on time means the value is 0
	$LTdelay_{i,s-1}$	The delay value of train i at station $s-1$, a positive value means delay, a negative value means early, on time means the value is 0
	$TTdelay_{i-1,s}$	The delay value of train $i-1$ at station s, a positive value means delay, a negative value means early, on time means the value is 0
	$LTdflag_{i,s-1}$	The delay sign of train i at station $s-1$
	$TTdflag_{i-1,s}$	The delay sign of train $i-1$ at station s
	$LTdiff_{s,s-1}$	The deviation of running time of train i from station $s-1$ to station s
	$TTdiff_{i,i-1}$	The deviation of running interval between train i and train $i-1$ at station s

The best delay classification scheme is an important basis for the classification of online data delay categories and has a direct impact on the subsequent prediction accuracy. Therefore, different classification schemes need to be compared to select the best scheme. For the DBSCAN algorithm, it is necessary to confirm the optimal classification number and the optimal combination of parameters (radius and minimum sample size). Table 2 shows the internal indicators that can complete the evaluation of the classification number.

Table 2 Cluster evaluation indicators

Index	Variable symbol	Correlation
Silhouette Coefficient	M_{1c}	Positive correlation
Calinski Harabasz Score	M_{2c}	Positive correlation
Davies-Bouldin Index	M_{3c}	Negative correlation

In order to compare the effectiveness of different delay classification schemes, each evaluation index is standardized. $M_{nc}^{'}$ represents the standardized result of the nth index. The range of the standardized indicators $M_{nc}^{'}$ is 0-1. For further comparison, the final weighted evaluation index is calculated by Eq. 2, the final weighted evaluation index is a negative correlation index.

$$M_c = \sum_{n=1}^{3} \alpha_n \cdot M_{nc}' \tag{2}$$

Herein: M_c is the final weighted index, α_n is the weight of the nth index, and $\sum_{n=1}^{3} \alpha_n = 1$, $0 \le \alpha_n \le 1$.

The best parameter combination will be determined by cross validation. Establish the clustering model under the corresponding parameter combination, and output the number of classifications, the number of abnormal points, and the number of sample points in each category. There are two principles for determining the optimal parameter combination. First, select a parameter combination with a small number of abnormal points. Second, choose a parameter combination with a relatively reasonable distribution of the number of sample points in each category.

4.2 Classification of delay types for online data based on KNN

After obtaining the best delay classification scheme, construct a data set containing delay type labels and complete the training of the delay classification model. The delay classification model will be completed using KNN algorithm. In KNN classification model, the K value has impact on the accuracy of prediction. Therefore, this paper uses the 10-fold cross validation method to determine the best K value. Fig. 4 shows the process of the classification method of delay type for online data.

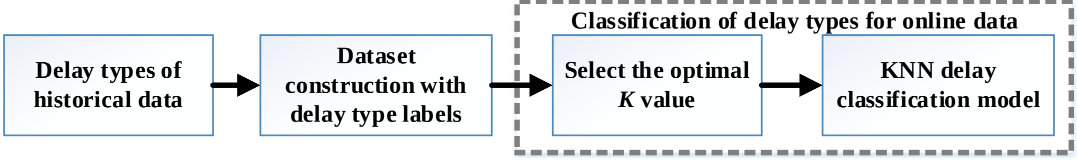


Fig. 4 The process of the classification of delay types for online data based on KNN

4.3 Feature extraction

The train operation status feature Z_i are variables related to the node $t_{i,s}$. First, the station information and arrival/departure information of the node should be considered. Secondly, variables related to $t_{i,s-1}$ and $t_{i-1,s}$ should be considered, including the delay value, the running interval between trains, and the train running time between stations, etc.

In order to better characterize the delay propagation relationship, this paper defines the delay propagation factor of each node. The calculation method is as shown in Eq. 3, where $factor_{i,s}$ is delay propagation factor of the node $t_{i,s}$, $D_{i,s}^n$ is the delay value of nth node in the delay propagation chain. In particular, for nodes that are early or punctual, the delay propagation factor is 0, and for nodes that are primary delay, the delay propagation factor is 1.

$$factor_{i,s} = \frac{D_{i,s}^n}{\sum_{n=1}^{N} D_{i,s}^n} \tag{3}$$

In the time event graph, each node has horizontal delay propagation and vertical delay propagation, the delay propagation factor is also different in the two directions, so there should be delay propagation factor in both horizontal and vertical directions corresponding to each node. The vertical delay propagation feature $P_{i,s-1}$ is related to the delay propagation of the train itself, so it should be related to the vertical delay propagation factor of $t_{i,s-1}$. The horizontal delay propagation feature $P_{i-1,s}$ is related to the delay propagation between trains, so it should be related to the horizontal delay propagation factor of $t_{i-1,s}$. Finally, the three type of input features of the GBDT model are shown in Table 3.

Table 3 Input feature set of GBDT model

Feature	Symbol	Meaning
Z_i	Sta_s	Station number of the sth station
	$A\_flag_{i,s}$	Arrival/departure sign of train i at station s
	$LTdelay_{i,s-1}$	The delay value of train i at station $s-1$, a positive value means delay, a negative value means early, on time means the value is 0
	$LTplan_{s,s-1}$	The planned running time of the train i from station $s-1$ to station s
	$TTdelay_{i-1,s}$	The delay value of the train $i-1$ at station s, a positive value means delay, a negative value means early, on time means the value is 0
	$TTplan_{i,i-1}$	The planned running interval between the train i and train $i-1$ at station s
$P_{i,s-1}$	$factor_{i,s-1}$	Delay propagation factor of $t_{i,s-1}$ (vertical)
$P_{i-1,s}$	$factor_{i-1,s}$	Delay spread factor of $t_{i-1,s}$ (horizontal)

5. Experiment and simulation

5.1 Data description

This paper uses the actual operation data of the 1H passenger express of the West Coast Main Line (WCML) railway in the United Kingdom. There are 37 stations on the route, and the time span is from June 1, 2017 to June 30, 2017, with a total of 47,693 records. The original data records information such as the train number, station number, actual or planned operating time, etc. Data from 75 % of the original data for model training and remaining 25 % of the original data for model testing.

5.2 Determine the type of delay

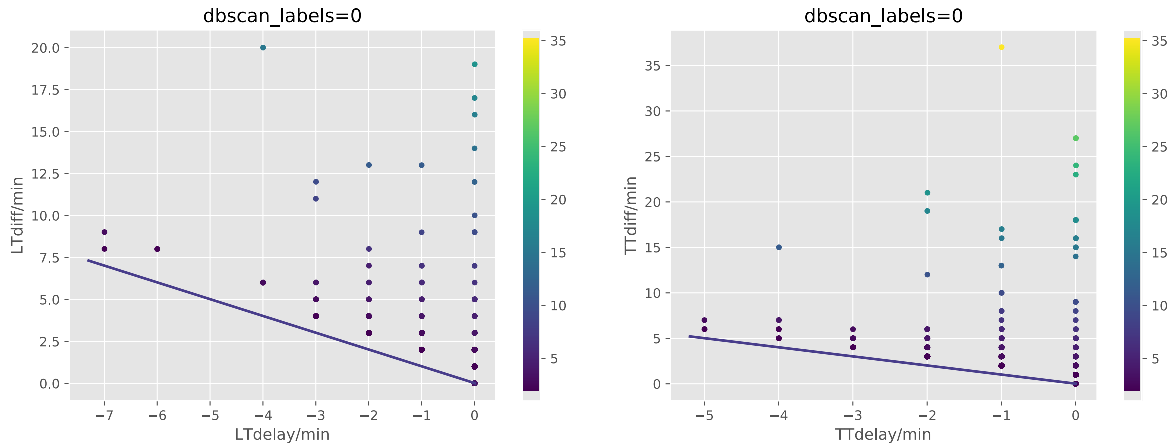
Use the weighted evaluation index M_c to determine the optimal classification number within the range of the classification number 2-10. The evaluation indicators under different classification numbers are shown in Table 4. When classification number is 9, M_c has a minimum value, so the best classification number is 9. In the range where the radius value is 0.05-3 and the minimum sample size value is 2-20, the DBSCAN model is constructed by cross validation, and get 21 groups of parameter combination results with the classification number of 9. According to the principle of selecting parameter combinations in section 4.1, the radius is 2.15, and the minimum sample size is 4.

According to the classification scheme, the delay type of each category is determined in a visual form. In the visual analysis, the analysis is carried out from the vertical and horizontal directions of the train delay propagation. Fig. 5 to Fig. 8 are the visual analysis diagrams of category 0-3, $Tdelay_{i,s}$ is represented by the color of the point in the scatter diagram.

The visualization of category 0 is shown in Fig. 5. In the vertical direction, the value of $TTdiff_{i,i-1}$ is greater than or equal to the value of $LTdelay_{i,s-1}$. In the horizontal direction, the value of $TTdiff_{i,i-1}$ is greater than or equal to the value of $TTdelay_{i-1,s}$. So the train was not affected by the delay propagation in both directions. The train was delayed during operation or was delayed at this station, so it was primary delay.

Table 4 The weighted evaluation index values under different classification numbers

Classification number c	M_c	Classification number c	M_c	Classification number c	M_c
2	0.682	5	0.257	8	0.195
3	0.839	6	0.274	9	0.015
4	0.315	7	0.209	10	0.046

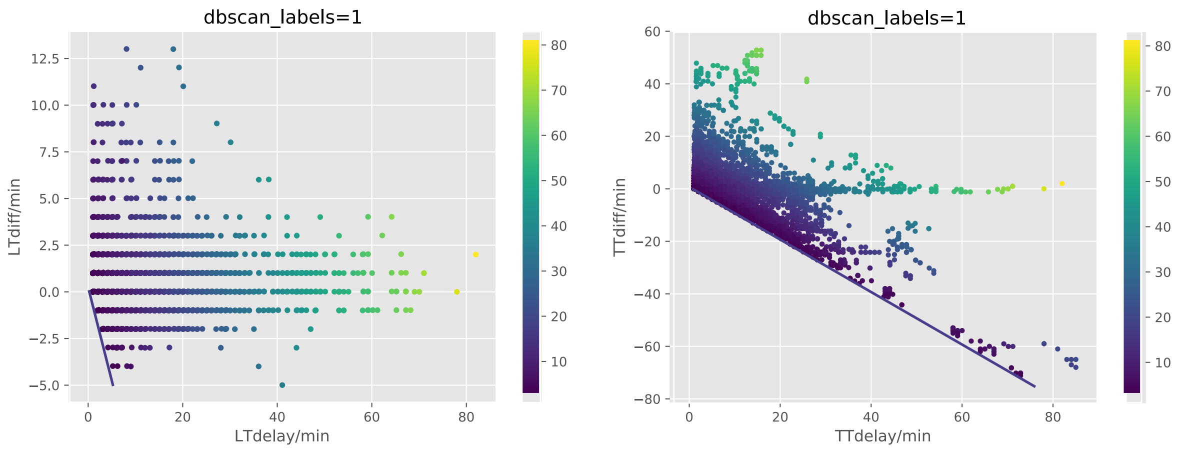


a) Scatter diagram of $LTdelay_{i,s-1}$ and $LTdiff_{s,s-1}$ b) Scatter diagram of $TTdelay_{i-1,s}$ and $TTdiff_{i,i-1}$

Fig. 5 Visualization of Category 0

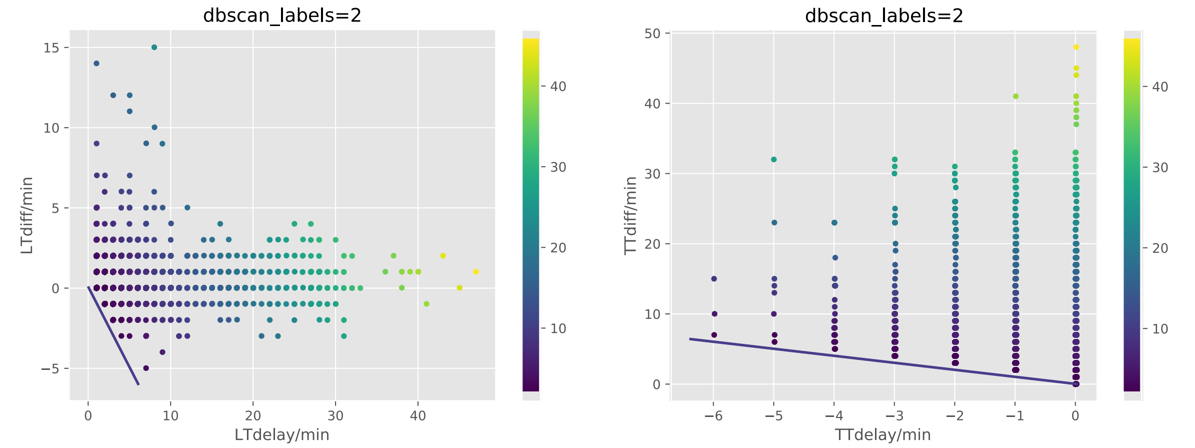
The visualization of category 1 is shown in Fig. 6. The delay of this train was affected by the delay propagation in both the horizontal and vertical directions, so it was secondary delay. The visual analysis results of categories 4, 5, 7, and 8 are the same as category 1. The visualization of the category 2 is shown in Fig. 7. The delay propagation occurred in the vertical direction, and the delay of the train at the previous station had an impact on this station, but there was no delay propagation in the horizontal direction, so it was secondary delay. The visual analysis results of the category 6 are the same as category 2. The visualization of category 3 is shown in Fig. 8. Contrary to category 2, category 3 is related to delays in horizontal directions, so it was secondary delay.

Finally, according to the characteristics of various types of delays in different directions, all data can be finally integrated into 4 types. The characteristics of the 4 types of delays are shown in Table 5, where delay type 1 is the primary delay, and the other three types are secondary delay. According to Table 5, the calculation methods of the delay propagation factors of the four types are different. Delay type 1 starts to propagate from the current node in the horizontal and vertical directions, and should be regarded as the primary delay in both directions. Delay type 2 are affected by the front nodes in both directions and should be calculated as secondary delay in both directions. Delay type 3 and delay type 4 are only affected by the front nodes in one direction, so it is regarded as secondary delay in one direction and the primary delay in the other direction.



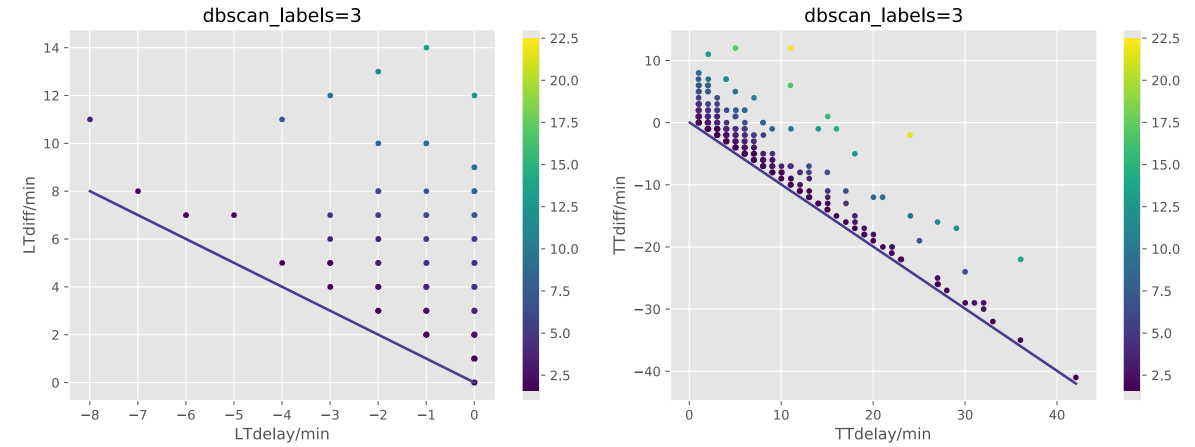
a) Scatter diagram of $LTdelay_{i,s-1}$ and $LTdiff_{s,s-1}$ b) Scatter diagram of $TTdelay_{i-1,s}$ and $TTdiff_{i,i-1}$

Fig. 6 Visualization of Category 1



a) Scatter diagram of $LTdelay_{i,s-1}$ and $LTdiff_{s,s-1}$ b) Scatter diagram of $TTdelay_{i-1,s}$ and $TTdiff_{i,i-1}$

Fig. 7 Visualization of Category 2



a) Scatter diagram of $LTdelay_{i,s-1}$ and $LTdiff_{s,s-1}$ b) Scatter diagram of $TTdelay_{i-1,s}$ and $TTdiff_{i,i-1}$

Fig. 8 Visualization of Category 3

Table 5 Results of four types of delays

Delay type	Category	Horizontal direction	Vertical direction	Result
1	0	No delay propagation	No delay propagation	Primary delay
2	1, 4, 5, 7, 8	Delay spread horizontally	Delay spread vertically	
3	2, 6	No delay propagation	Delay spread vertically	Secondary delay
4	3	Delay spread horizontally	No delay propagation	

5.3 Model performance

On the basis of determining the delay type, construct a training set containing the label of the delay type, and determine the K value of the KNN classification model to be 4. In order to verify the performance of GBDT-PF model, this paper will compare the random forest (RF), support vector regression (SVR) and multilayer perceptron (MLP) that are widely used in train delay prediction, and will also verify the importance of the delay propagation feature. The optimal parameter combination of four models is shown in Table 6. This paper uses three evaluation indicators include root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R2) to evaluate the parameter combination. Table 7 shows the RMSE, MAE and R2 of each model.

Table 6 The optimal parameter combination of four models

Model	Parameter	The value under different feature combinations	
		Z_i	$Z_i, P_{i,s-1}, P_{i-1,s}$
GBDT	learning_rate	0.49	0.06
	n_estimators	96	91
	min_samples_split	300	284
	min_samples_leaf	2	3
	max_depth	5	17
	max_feature	5	3
RF	n_estimators	80	90
	max_features	4	6
	max_depth	7	9
SVR	C	3.2	2.1
	loss	epsilon_insensitive	epsilon_insensitive
MLP	hidden_layer_sizes	(20,20,20)	(80,80,80)

Table 7 The index values of each model under different feature combinations

Model	Input feature	RMSE	MAE	R2
RF	Z_i	1.53354	0.51600	0.94404
	$Z_i, P_{i,s-1}, P_{i-1,s}$	1.44952	0.44150	0.95000
SVR	Z_i	1.67973	0.79100	0.93286
	$Z_i, P_{i,s-1}, P_{i-1,s}$	1.58231	0.62100	0.94042
MLP	Z_i	1.69797	0.64000	0.93140
	$Z_i, P_{i,s-1}, P_{i-1,s}$	1.66474	0.60150	0.93405
GBDT	Z_i	1.40097	0.40600	0.95330
	$Z_i, P_{i,s-1}, P_{i-1,s}$	1.35860	0.37400	0.95608

After adding the delay propagation feature, the index value of each model is optimized. Therefore, considering the impact of the delay propagation in the delay prediction can improve the prediction accuracy. Among the four models, the GBDT model with delay propagation feature performs better on three indicators. Fig. 9 shows the distribution of the prediction errors of each model.

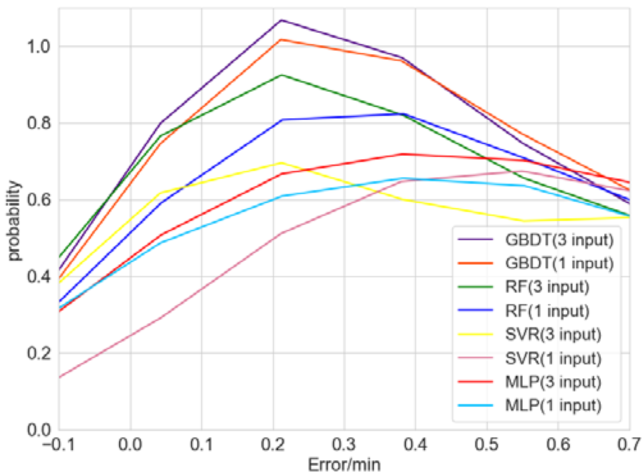


Fig. 9 Error distribution of each models under different feature combinations

It can be seen from Fig. 9 that the peak value of the model error distribution curve containing the delay propagation feature is closer to the vertical axis, indicating that the overall error is smaller. The error distribution curve of the GBDT model with delay propagation feature is closest to the vertical axis in all models, so its overall error is the smallest and the prediction accuracy is higher.

5.4 Model simulation

In order to display the prediction results of the delay prediction model more intuitively, this paper uses the PYQT5 package to complete the simulation of the model on the Pycharm software. The simulation program design process completed by PYQT5 is shown in Fig. 10.

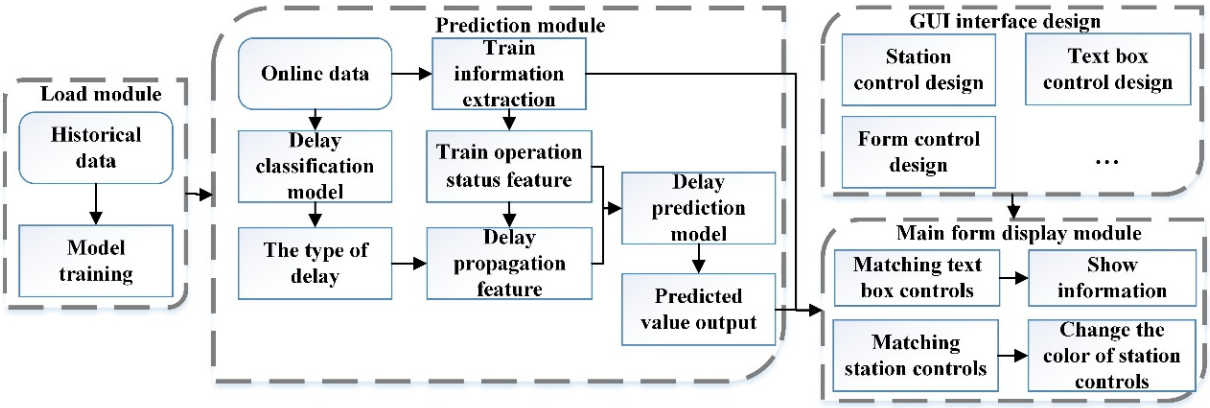


Fig. 10 The design process of simulation program for train delay prediction based on PYQT5

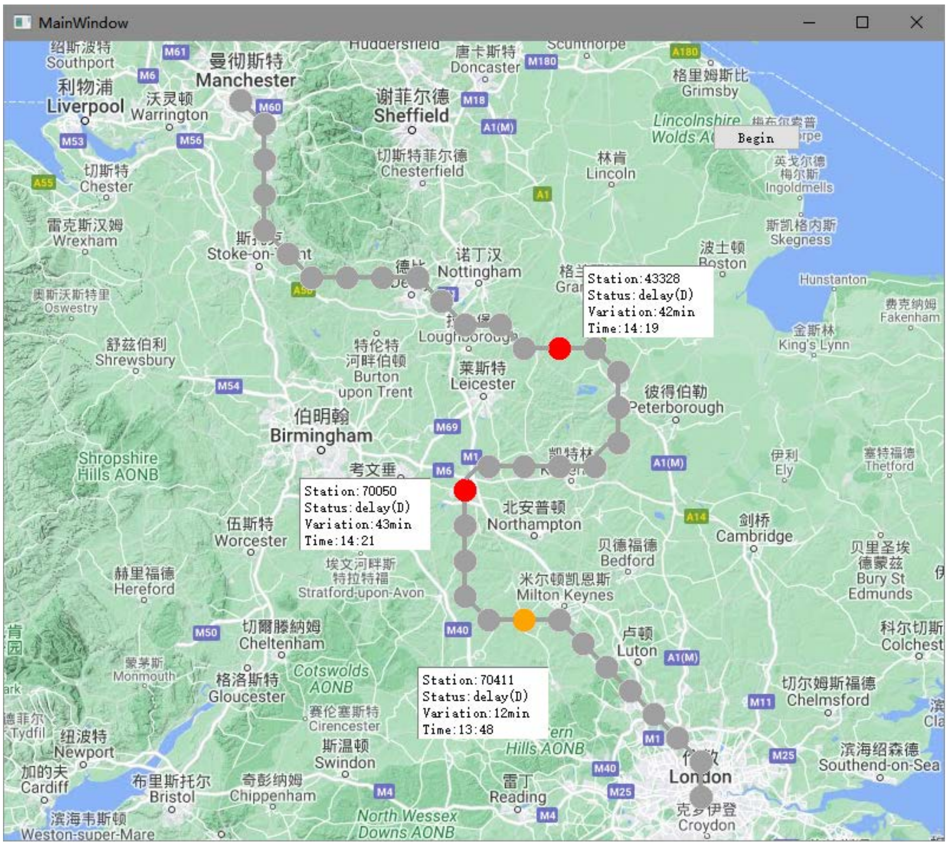


Fig. 11 Train delay prediction simulation interface

There are two main functions of the simulation program: (1) It can display the train information, station information, departure or arrival time in real time. At the same time, different colors are used to indicate the current degree of delay. Green means the train is on time, blue means the train is early, yellow means the train is delayed within 5 minutes, orange means 5-15 minutes delay, and red means more than 15 minutes delay; (2) According to train operation data and prediction model, display the operation status of each train on the line dynamically. According to the simulation interface, the prediction results can be viewed in real time, and the delay propagation phenomenon can be observed. Fig. 11 shows the operation of the three delayed trains on the line through simulation interface.

6. Conclusion

This paper proposes a GBDT-PF model that considers the delay propagation feature. The effectiveness of the method is evaluated by taking the train operation data of the British WCML line as an example, and the following conclusions are drawn:

- Based on the characteristics of primary delay and secondary delay in the delay propagation, using DBSCAN algorithm to design a clustering method of delay types for historical data, through this method, the delays can be finally divided into four categories. The four types of delays have obvious characteristics in the vertical and horizontal direction. And according to the best delay classification scheme, the KNN algorithm is used to design the classification method for online data to identify the type of delay.
- Based on the results of the identification of delay types, the delay propagation relationship is quantified by the delay propagation factor and used as the input feature of the GBDT model. According to the experimental comparison results, when predicting train delays, considering the delay propagation feature can further improve the prediction accuracy.

With the development of railway informatization, based on the comprehensive collection of actual train operation data, the dispatching and commanding of railway trains will also be more intelligent. The delay prediction model proposed in this paper can provide delay prediction data for intelligent dispatch and make the dispatching and command work more efficient.

Acknowledgement

This work was supported by Sichuan Science and Technology Program (2021YJ0070).

References

[1] Huang, P., Peng, Q., Wen, C., Yang, Y. (2018). Random forest prediction model for Wuhan-Guangzhou HSR primary train delays recovery, *Journal of the China Railway Society*, Vol. 40, No. 7, 1-9.

[2] Wen, C., Li, Z., Huang, P., Tian, R., Mou, W., Li, L. (2019). Progress and perspective of data-driven train delay propagation, *China Safety Science Journal*, Vol. 29, No. S2, 1-9, doi: 10.16265/j.cnki.issn1003-3033.2019.S2.001.

[3] Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D. (2018). Train delay prediction systems: A big data analytics perspective, *Big Data Research*, Vol. 11, 54-64, doi: 10.1016/j.bdr.2017.05.002.

[4] Wang, P., Zhang, Q.-P. (2019). Train delay analysis and prediction based on big data fusion, *Transportation Safety and Environment*, Vol. 1, No. 1, 79-88, doi: 10.1093/tse/tdy001.

[5] Shi, R., Xu, X., Li, J., Li, Y. (2021). Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization, *Applied Soft Computing*, Vol. 109, Article No. 107538, doi: 10.1016/j.asoc.2021.107538.

[6] Nair, R., Hoang, T.L., Laumanns, M., Chen, B., Cogill, R., Szabó, J., Walter, T. (2019). An ensemble prediction model for train delays, *Transportation Research Part C: Emerging Technologies*, Vol. 104, 196-209, doi: 10.1016/j.trc.2019.04.026.

[7] Li, Z.-C., Wen, C., Hu, R., Xu, C., Huang, P., Jiang, X. (2020). Near-term train delay prediction in the Dutch railways network, *International Journal of Rail Transportation*, Vol. 9, No. 6, 520-539, doi: 10.1080/23248378.2020.1843194.

[8] Huang, P., Wen, C., Fu, L., Lessan, J., Jiang, C., Peng, Q., Xu, X. (2020). Modeling train operation as sequences: A study of delay prediction with operation and weather data, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 141, Article No. 102022, doi: 10.1016/j.tre.2020.102022.

[9] Gao, B., Ou, D., Dong, D., Wu, Y. (2020). A data-driven two-stage prediction model for train primary-delay recovery time, *International Journal of Software Engineering & Knowledge Engineering*, Vol. 30, No. 7, 921-940, doi: 10.1142/S0218194020400124.

[10] Tang, Y., Xu, C., Wen, C., Li, Z., Song, S. (2019). Support vector regression models for delay time predicting considering high-speed rail facility failure, *China Safety Science Journal*, Vol. 29, No. S2, 18-23, doi: 10.16265/j.cnki.issn1003-3033.2019.S2.003.

[11] Zhang, Q., Chen, F., Zhang, T., Yuan, Z.M. (2019). Intelligent prediction and characteristic recognition for joint delay of high speed railway trains, *Acta Automatica Sinica*, Vol. 45, No. 12, 2251-2259, doi: 10.16383/j.aas.c190188.

[12] Hu, R., Xu, C., Feng, Y., Wen, C., Wang, Q. (2019). Prediction of different types of train delay of Guangzhou-Shenzhen high-speed railway, *China Safety Science Journal*, Vol. 29, No. S2, 181-186, doi: 10.16265/j.cnki.issn1003-3033.2019.S2.030.

[13] Zeng, Y., Chen, F., Jin, B. (2019). A prediction model for timetable delays in dispatching area using neural network, *Railway Standard Design*, Vol. 63, No. 3, 148-153, doi: 10.13238/j.issn.1004-2954.201812160002.

[14] Milinković, S., Marković, M., Vesković, S., Ivić, M., Pavlović, N. (2013). A fuzzy Petri net model to estimate train delays, *Simulation Modelling Practice and Theory*, Vol. 33, 144-157, doi: 10.1016/j.simpat.2012.12.005.

[15] Lessan, J., Fu, L., Wen, C. (2019). A hybrid Bayesian network model for predicting delays in train operations, *Computers & Industrial Engineering*, Vol. 127, 1214-1222, doi: 10.1016/j.cie.2018.03.017.

[16] Pullagura, L., Katiravan, J. (2019). Train delay prediction using machine learning, *International Journal of Engineering and Advanced Technology (IJEAT),* Vol. 9, No. 2, 1312-1315, doi: 10.35940/ijeat.A2088.129219.

[17] Huang, P., Wen, C., Fu, L., Peng, Q., Tang, Y. (2020). A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems, *Information Sciences*, Vol. 516, 234-253, doi: 10.1016/j.ins.2019.12.053.

[18] Hansen, I.A., Goverde, R.M.P., van der Meer, D.J. (2010). Online train delay recognition and running time prediction, In: *Proceedings of 13th International IEEE Conference on Intelligent Transportation Systems,* Funchal, Portugal, 1783-1788, doi: 10.1109/ITSC.2010.5625081.

APEM journal
Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 297–306
https://doi.org/10.14743/apem2021.3.401
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

A smart Warehouse 4.0 approach for the pallet management using machine vision and Internet of Things (IoT): A real industrial case study

Vukićević, A.[a], Mladineo, M.[b], Banduka, N.[a,b], Mačužić, I.[a,*]

[a]Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia
[b]Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split, Croatia

A B S T R A C T

Printing companies are commonly SMEs with high flow of materials, which management could be significantly improved through the digitalization. In this study we propose a smart Warehouse 4.0 solution by using QR code, open-source software tools for machine vision and conventional surveillance equipment. Although there have been concerns regarding the usage of QR in logistics, it has shown to be suitable for the particular use-case as pallets are static in the inter-warehouse. The reliability of reading of QR codes was achieved by using multiple IP cameras, so that sub-optimal view angle or light reflection is compensated with alternative views. Since surveillance technology and machine vision are constantly evolving and becoming more affordable, we report that more attention needs to be invested into their adaptation to fit the needs and budgets of SMEs, which are the industrial cornerstone in the most developed countries. The demo of proposed solution is available on the public repository *https://github.com/ArsoVukicevic/PalletManagement*.

A R T I C L E I N F O

Keywords:
Smart manufacturing;
SME;
Industry 4.0;
Logistics 4.0;
Warehousing 4.0;
Pallet management;
Machine vision;
Internet of Things (IoT);
QR code

**Corresponding author:*
ivanm@kg.ac.rs
(Mačužić, I.)

Article history:
Received 30 May 2020
Revised 15 September 2021
Accepted 17 September 2021

1. Introduction

All manufacturing-oriented companies, both SMEs and large corporations, are on a daily basis faced with the problem of managing a large number of different articles: which may vary from various raw materials to components and parts. Thus, the success of each business depends on the continuous and fluent flow of the parts in the supply chain [1]. However, the real-world situation is usually opposite, since there are huge problems in establishing of the reliable supply chain, leading to increase on inventory levels to fight against fluctuations and other disturbances. Modern management strategies, including Lean and World class manufacturing (WCM), define an unwanted level of inventory (raw material, work in progress, or finished products) as waste since they occupy working capital [2]. So, if there is no successful warehouse management in every company in the supply chain, all these issues cannot be properly addressed.

The new industrial platform – Industry 4.0 [3-6], aims to ease and automate the tracking and management of material flow, thus improving it on the global level (supply chain management) and on the local level (warehouse management). New Warehousing 4.0 solutions are based on the Smart Products. Therefore, they need to be identifiable at any time – which requires sophisticated technology [3]. The new approaches are based on the application of QR (Quick Response) code and RFID (radio-frequency identification) tags, and other I4.0 technologies such as IoT, cloud and data mining [7, 8]. Tracking of unique products allows lifecycle management of single-item customized products and optimization of their production flow [9].

In general, solutions based on QR and RFID have been the most widely used since they are affordable and easy to use. These are important aspects, since the companies in many countries are lagging behind with Industry 4.0 technology adoption [10]. Although there are concerns regarding the security of QR code, it has been widely used in marketing (opening web-page URL written in QR code) and/or product tracking [11]. On the other hand, RFID technology uses RFID tags for storing information, and corresponding antennas to read and write data to the tag [12] with possibility to use data encryption. Although the RFID technology is more attractive, since much more data can be stored in the tag, there is a serious concern about bad influence of the RFID technology on the human body. It is a non-ionizing type of radiation, but some researches show that it could have a negative impact on the human body in a long-term period [13, 14]. Therefore, the range of antennas used in the industrial RFID solutions has been limited to below 0.1 m, although the range could be more than 1 m [15]. It means that RFID tags cannot be read from the distance, but from the vicinity, like the classic one-dimensional barcode. The possibility to read QR code from the distance by using high-resolution cameras becomes the advantage of the QR code over the RFID technology.

In 2012, Qian *et al*. investigated the possibilities of tracking two-dimensional QR codes in the food industry, and compared them with the RFID technology [16]. Their evaluation found that the implementation of mentioned solutions increased overall cost for 17.2 % but increased efficiency and sales up to 32.5 %. Liang *et al*. proposed usage of QR code for automatic separation and identification of equipment [17]. In the recent study from 2020, Liu *et al*. proposed a system based on IoT that enables remote reading of barcodes for the purpose of implementing "smart cities" concept [18]. Regarding the industry practice, QR codes are widely used for employee authentication[1] – however, it is not widely used in the logistics. This is because the QR technology has not been shown as reliable in cases where it needs to frequently read moving objects (i.e. baggage check in airports, markets etc.). The leading manufacturer of industrial solutions for the barcode-based tracking of pallets is the Cognex[2] company, and their solutions are based on using industrial cameras and accompanying information systems. However, limited budgets and complexity of such solutions is the major obstacle for their wider usage among SMEs. On the other side, development of dedicated solutions was limited with costs of professional industrial cameras, which are needed for ensuring high quality of images for reading QR codes from distance.

The SME from the printing industry, a use-case in this study, was considered as the representative case because there are literally hundreds of different variations (formats, weights, colors, origins, fiber orientations, surfaces, printing characteristics, etc.) of paper in the production process and supporting materials that are used daily [19]. Particularly, this study focuses on covering the needs of SMEs since they generate the most workplaces and GDPs of developed countries - thus covering their needs may have considerable impact on the economy and society [20, 21].

The starting assumption of this study is that above-mentioned barriers for QR code application are vanishing with the technological progress, since even a standard IP camera delivers images with 8+ megapixels. And when it comes to the transport of materials within a company, it is important to emphasize that forklifts and pallets still represent the gold standard [1], like in this use-case. However, to the best of our knowledge, there is no scientific study, nor publically available solution, that assesses their potential and usability for a particular industrial purpose of tracking pallets in the inter-warehouse of a printing company. Accordingly, the aim of this study was to investigate how QR code, machine vision and IP cameras could be adapted for improving pallet management in a representative SME printing company.

[1] *http://www.mytimestation.com* [22]
[2] *https://www.cognex.com/* [23]

2. Materials and methods

2.1 Machine vision

Computer vision (CV) is an emerging scientific field that falls under the umbrella of Machine learning (ML) and Artificial intelligence (AI). It groups algorithms that performs decision making on the basis of observed visual inputs. CV is primarily focused on making higher-level decisions by processing data that may not be only 2D images, but also point clouds, meshes, videos, etc. On the other side, Machine vision (MV) is nowadays identified as a subfield of CV that refers to the use of CV and image processing techniques in industry setups (which means use of dedicated industry cameras, lenses, PCs, etc.) for making decisions (commonly real-time). Typical examples of the use of MV in manufacturing industry are dimensional inspection [24] and defects inspection [25].

2.2 Internet of Things

The term Internet of Things (IoT) refers to the use of devices connected to the internet, with the purpose to measure and collect data, or control remote devices. Similarly, with the increase of the use of IoT, there is appearing Industrial IoT (IIoT) as a separate scientific topic. As two key underpinning technological pillars of the Industry 4.0, the use of CV and MV together with IoT shown a high potential for solving wide range of problems. Particularly, with the advances of networking and constant increase of IP cameras affordability and quality of images that can provide in real-time. Accordingly, they have appearing as possible replacement for more expensive industrial cameras – especially for task such are reading QR code and tracking material flow in warehouses.

2.3 QR code technology

Understanding the key concepts of QR technology is necessary for its successful application in industry, since a developed solution needs to meet related industry standards[3]. The QR (Quick Response) was invented back in 1994. by Japanese company Denso Wave[4], and it is composed of parts shown in Fig. 1 The gray area indicates a clear zone with no data. Square elements in corners (bottom left, top left and right) are used for detection of scale and orientation of the code. Different colors show specific zones that contain version, format, and QR code information. There are about forty versions of QR code, and the most commonly used are versions 1 and 2, which have 21 × 21 and 25 × 25 modules, respectively (dimensions determine amount of data that could be written). The data zone also contains elements necessary for correcting errors on printed codes, which enables reading damaged QR codes. There are four code correction levels[5]: L (7 %), M (15 %), Q (25 %), and H (30 %). In this study, we used version two with the M correction level.

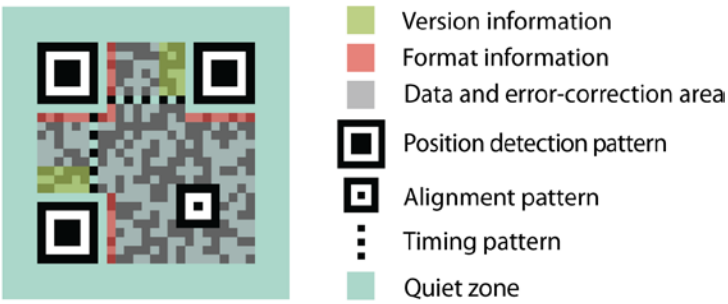


Fig. 1 QR code composition as documented by the Denso company

[3] *https://www.iso.org/standard/62021.html* (ISO/IEC 18004:2015) [26].
[4] It was initially invented for the purpose of tracking manufactured parts in the automotive industry.
[5] It should be taken into account that error correction levels are related to the magnitude of the QR code (a greater degree of correction requires larger dimensions of the code).

2.4 Used equipment

The development of the procedure was completely done in lab conditions, by using minimized QR codes observed with multiple conventional USB web cameras. Particularly, this was possible because technologies utilized in this study have robust interface for accessing various types of cameras with minimal change of software code. The equipment used for the deployment into the real-industry conditions included a conventional PC, network switch and NVR device connected with 8MP Dahua IP cameras that were mounted on the warehouse roof construction.

3. A case study: Proposed pallet management system in real industrial environment

Photo from the use-case factory is shown in Fig. 2 The current practice for managing the material flow in inter-warehouse is based on the standard principles of work orders in the printing industry [19], [27-29]. Trained employees control material disposal and storage zones, and the space is used in accordance with current needs and requirements determined by the weekly work plan and customers' requirements. Due to the dynamics of the business, and the company growth, there is an increasing need for a system able to ease the control of pallet flow and prevent frequent delays. Particularly, the continuous and increasing production often results in accumulated inter-resources/products, significant space congestion and waste of time that employees spent to search for missing pallets and parts necessary for the initiation of further manufacturing processes. Although the warehouse space is graphically coloured, it rarely easies employees' efforts because of the previously explained reasons.

3.1 User requirements

From the SME management standpoint, the disadvantages of the current practice are:

- Due to the high frequency and overload, forklift drivers often do not comply FIFO rules – which results with misallocated pallets.
- Pallets that are misallocated significantly increase the time needed to afterwards find and sort the pallets available in the warehouse. For the company, it is very important to be able to quickly identify complete work orders and the location of their pallets (tabs).
- Failure to accurately and efficiently inventory all available pallets in the storage causes delays in subsequent manufacturing operations. Process engineers have reported that the delay could be up to three hours - causing significant losses in the profit and productivity.
- The solution needs to be easy to use and accessed by both management and employees on the site. It should be able to deliver a fast report, which contains documented current status and alignment of materials (pallets) through the production space.

Fig. 2 A typical temporary or intra-warehouse of a printing SME. After finishing one operation, the tabs are stacked on pallets until they are requested by a following manufacturing process

3.2 Pallet management by using QR code, machine vision and IP cameras

Concept of the proposed solution is illustrated in Fig. 3. The solution was released as a series of modules, which do particular tasks: 1) Image acquisition, 2) QR codes detection, 3) User interaction (GUI), and 4) Reporting.

The image acquisition was carried out with conventional IP cameras that were fixed on a roof construction (Fig. 4c). Key components of the surveillance[6] system are IP cameras, network video recorder and central management software. The central PC-server runs the Python application (the GUI was developed by using the Qt5 library), which processes incoming images on user request. Beside the standard Python libraries for data structures and numeric, OpenCV (image acquisition and processing) and python-docx library (generation of MS Word documents) were used, together with the QR code library for processing QR codes. Detailed architecture and UML workflow diagram of the solution are given in Fig. 4 In Fig. 5, we present the captures of our implementation in an SME that annually processes ~40 million tabs of paper.

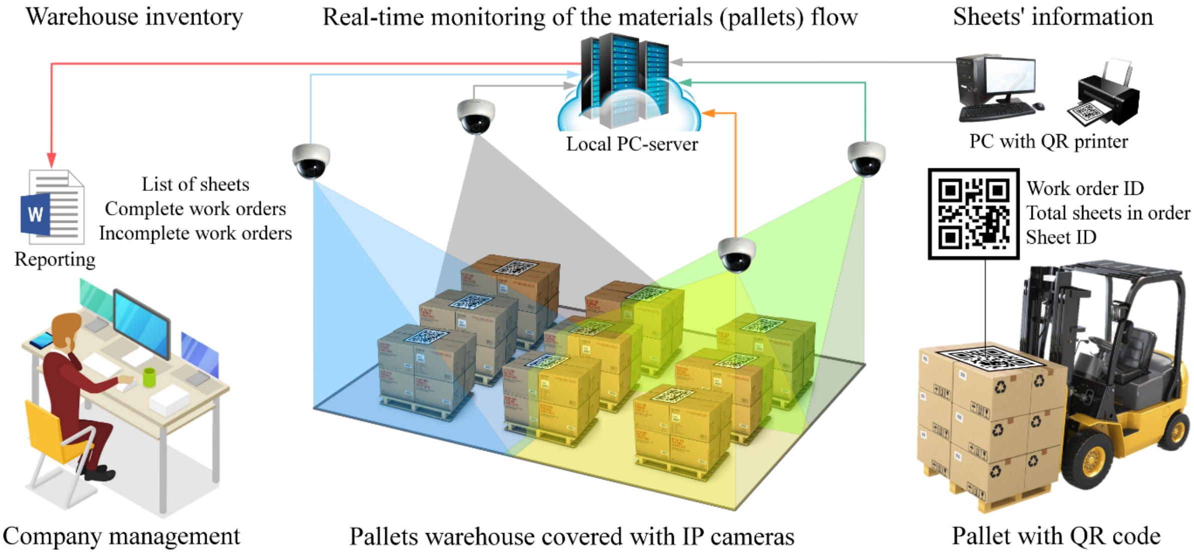


Fig. 3 Concept of the proposed solution

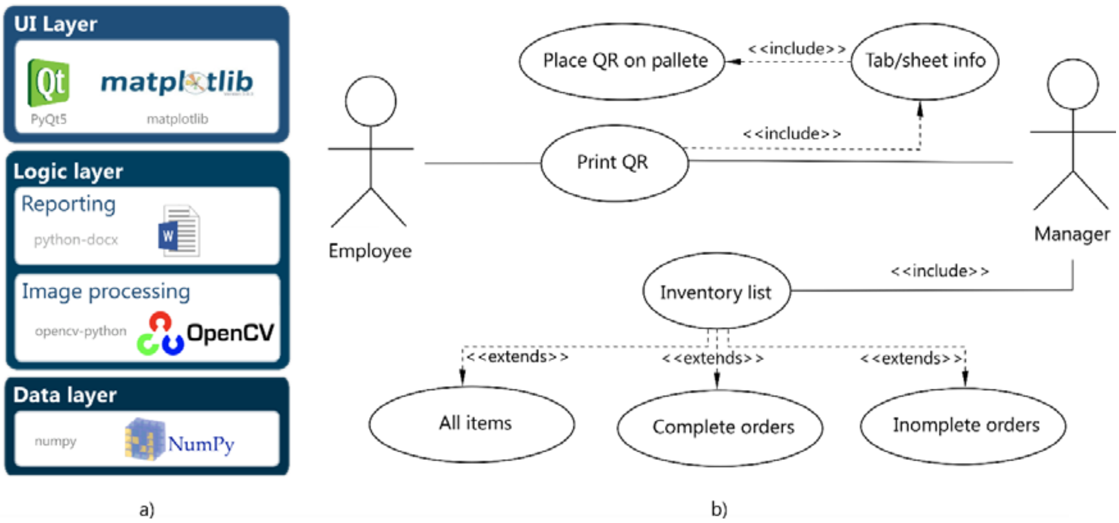


Fig. 4 Software architecture and UML Workflow diagram of the proposed solution

[6] For the purpose of this study, we used DAHUA IPC-HFW2831TP-ZS 8MP WDR IR Bullet IP Camera (4x), with DAHUA PFS3010-8ET-96 8 port Fast Ethernet PoE switch. Host PC had CPU 1151 INTEL Core i3-8100 3.6 GHz 6 MB BOX, results were visualized in HISENSE 40" H40B5100 LED Full HD digital LCD TV and printed on Printer HP LaserJet Pro M102a.

The starting point of the workflow is the moment when a manager assigns a work order to an employee. Orders physically represent a series of tabs / sheets transported on pallets with forklifts (Fig. 4, right). For each incoming tab, one has to print an appropriate QR code by using the proposed software application. The generated QR code contains the following information that determine one tab: 1) work order ID, 2) total tabs within the order, and 2) tab ID. The adapted format of the string data written in QR code is "XXXX ,YY/ZZ": the XXXX indicates the order ID (maximum is 9999 annually), YY indicates the total number of tabs in the work order (maximum 99) and ZZ indicates the tab ID. Below the QR code (which is not readable to humans), we print readable information and time of the code generation, which are fused on a single file and printed on the printer connected to the local network (Fig. 5b). The sample QR code placed on the pallet is shown in Fig. 5d, while the print format is shown in Fig. 5e. The left side of the screen is reserved for user commands (printing QR, querying detected tabs etc.). Visualization of results is done in the middle of the screen, while on the right side are lists of the inventory result. The user is allowed to query the inventory list with respect to orders (complete, incomplete) and tabs. In order to enable fast and intuitive inspection, we colored items found in the lists – so that the red items/orders indicate error, yellow items indicate incomplete orders and green indicate complete orders. On this way, employees could easily spot complete orders, and initiate the accompanying manufacturing process that waits for these pallets to be transported through the manufacturing hall. Another supported functionality is the generation of MS Word reports, which include a list of all orders and corresponding tabs (grouped on complete and incomplete). In this way, the company could document and track the pallets flow and management through time - with the aim to spot bottlenecks and make improvements.

Fig. 5 Preview of the implemented solution: a) Large screen and server (including PC, switch and NVR device) placed in the protective box; b) Printer placed in the protective box; c) Dahua 8MP IP camera mounted on the roof construction; d) QR code placed on the top of the palate; e) QR code format

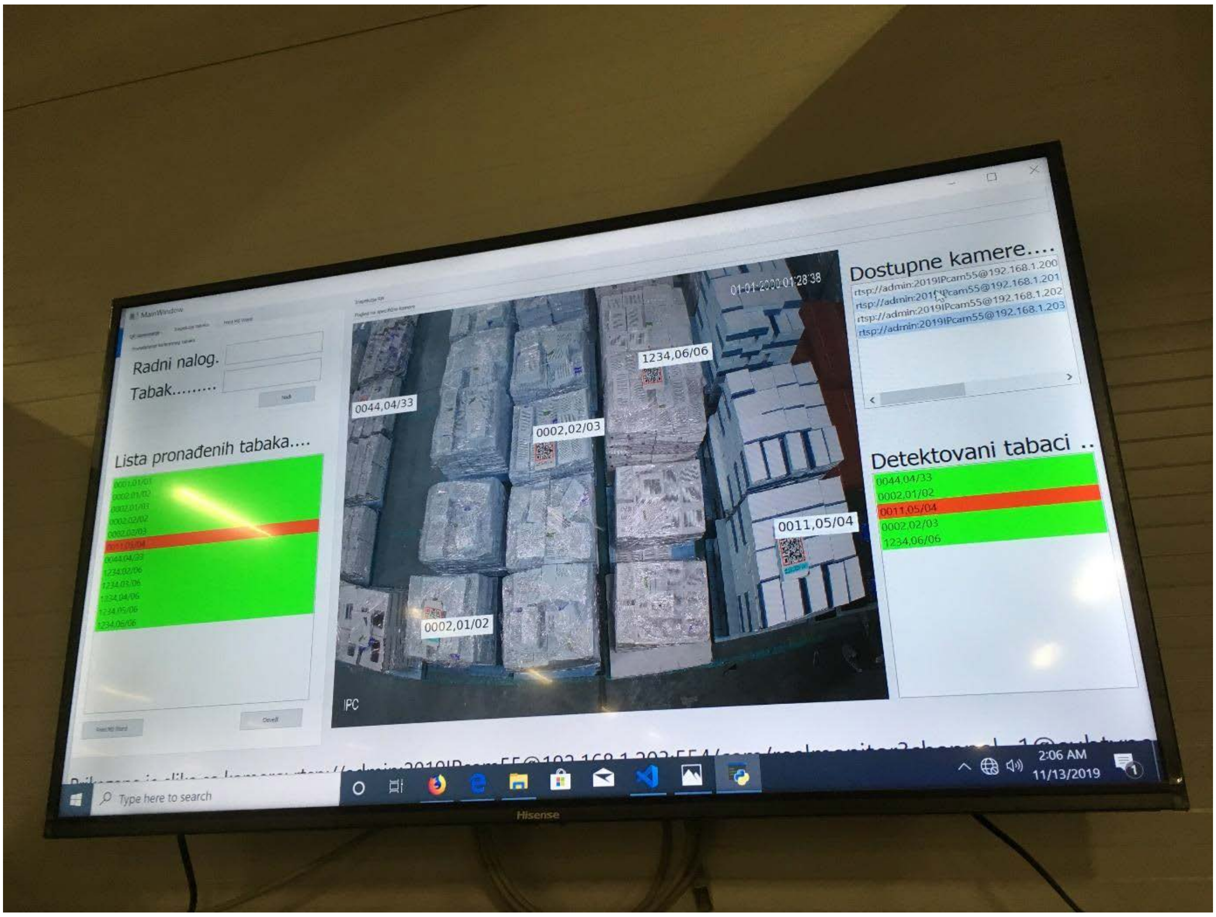


Fig. 6 Capture of the system usage in industry conditions. The upper left are commands for querying and printing of QR codes, while the below are lists of tabs (a user is allowed to view all tabs or only tabs that belong to particular work orders). In the middle are visualized positions of tabs that belong to a particular work order. On the right side are a list of cameras and a list of tabs visible on the particular camera

4. Discussion

Materials Management Systems (MMS), independently or as part of Enterprise Resources Planning (ERP) systems have been one of the interesting and promising directions for the development of supported hardware/software supply chain platforms in recent years [30]. Although the available ERP solutions are numerous and varied (from complex to simple ones), there is still a significant lack of ERP modules or independent systems that can enable (near) real-time monitoring and monitoring of material status and flow at the store level (warehouses, warehouse monitoring and inventory tracking). For this reason, the need for research and practical work to address the problem identified is justified and it could lead to results that can make progress for both the industry and the scientific community. Since the most of printing companies are SMEs, investment into complex or expensive commercial solution often represents a major obstacle towards their digitalization. Accordingly, the proposed solution was developed with aim to avoid these obstacles (the overall hardware installation costs were about 2000 EUR), and it represents so-called "low-cost automation" [31, 32]. Particularly, we emphasize the needs of SMEs because they generate the most of the GDP and employment opportunities in developed countries [20]. Therefore, although there are more robust and general-purpose ERP solutions – still a lot of effort needs to be invested into development of dedicated solutions for specific problems. To sum up, dedicated alternatives, such as the proposed solution, may represent a valuable improvement of the current pallet management practice and may result with considerable impact in SME industry sectors. Accordingly, we made the proposed solution available on the public repository: *https://github.com/ArsoVukicevic/PalletManagement* [33].

In this study, the reliability problem of QR readers [34] was solved with the usage of multiple IP cameras. Briefly, if one camera fails to detect a pallet (due to a suboptimal viewing angle, lighting etc.), another IP camera commonly succeeds. On this way, chances that some pallet may be omitted are drastically decreased. We emphasize that the solution was realized by using four 8-megapixel Dahua IP cameras, and we do not recommend to use low-resolution or lower-tier brands. The aforementioned pixels and the number of cameras have shown to be quite sufficient to cover a space of 10 x 20 meters from the height of about 4.5 meters (which corresponds to the height of the roof support in the manufacturing hall). Furthermore, we preferred to print QR codes on a conventional PC printer, in A4 format, which is widely used in offices. Although we have not experimented with smaller paper sizes, the assumption is that a professional QR printer could be also used - but, in this way, one would increase the cost of equipment and software development. The major limitation of the proposed solution is that the reading QR code is its sensitivity to folding – which we solved by cardboard carrier. The remaining concern of reading false QR codes (e.g. that could appear on products or parts available in factory) or misuse and misplace of QRs is unlikely to happen in industry environment by company employees. Regarding the impact of the developed system on the logistics in the considered company, as the inventory list is done in terms of a few seconds – the achieved speed is incomparably higher than the possibility of a manual search and inspection of the warehouse by employees. In particular, we remind that, based on the employee experience and reports, manual searching for the missing pallets can take up to three hours. For this reason, the advantage of the digitized system is obvious and there was no need for statistical analysis of the performance improvement. Finally, the solution is easily adaptable to various environments – as the only adjustment needed is mounting of conventional IP cameras. The further work on this topic may be regarded to improvement of security, trackability and integrability into existing ICT platform, for which various technologies could be applied including blockchain [35].

5. Conclusion

Beside technological advancements that bring Industry 4.0, the flow of products and raw materials within companies is still underpinned with forklifts and pallets. Since management of pallets in inter-warehouse still depends on human factors, many companies are faced with delays caused by operator's inability to timely manage large amounts of pallets. As a representative use-case, we considered a SME company from the printing industry since there are literally hundreds of different variations of paper (formats, weights, colors etc.) in the production process. Although there are available commercial warehouse barcode-based solutions for tracking pallets, complexity and cost of such solutions represent major obstacles for their wider usage by SMEs (which budgets are limited). As an alternative, we assessed the low-cost solution for the pallet management with QR code, machine vision and IP cameras (demo is available at *https://github.com/ArsoVukicevic/PalletManagement* [33]). The compact solution was developed by using free open-source software libraries and conventional surveillance equipment. Although there have been concerns regarding usage of QR code-based solutions for tracking, we report that it is suitable for the particular purpose of warehouse inventory since pallets are static. Reliability of the solution was ensured by using multiple IP cameras, which ensures that if one camera fails to detect QR code another one will compensate for it. The recent evolution of IP cameras (which now have 5+ megapixels) made them affordable and efficient tools for reading QR codes from larger distances. Thus, we conclude that more attention and effort need to be invested into investigation and adaptation of widely available technologies that could fit the needs of SMEs.

Acknowledgement

This study represents the results of the project Advanced system for material tracking and waste reduction funded by European Bank for Reconstruction and Development (EBRD) through the Green Innovation Vouchers program.

References

[1] Roy, D., Carrano, A.L., Pazour, J.A., Gupta, A. (2016). Cost-effective pallet management strategies, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 93, 358-371, doi: 10.1016/j.tre.2016.06.005.

[2] Costantino, N., Dotoli, M., Falagario, M., Fanti, M.P., Mangini, A.M. (2012). A model for supply management of agile manufacturing supply chains, *International Journal of Production Economics*, Vol. 135, No. 1, 451-457, doi: 10.1016/j.ijpe.2011.08.021.

[3] Kagermann, H., Wahlster, W., Helbig, J. (2013). *Recommendations for implementing the strategic initiative Industrie 4.0*, Heilmeyer und Sernau, Berlin, Germany.

[4] Jayaraman, V., Ross, A.D., Agarwal, A. (2008). Role of information technology and collaboration in reverse logistics supply chains, *International Journal of Logistics Research and Applications*, Vol. 11, No. 6, 409-425, doi. 10.1080/13675560701694499.

[5] Tang, C.S., Veelenturf, L.P. (2019). The strategic role of logistics in the Industry 4.0 era, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 129, 1-11, doi: 10.1016/j.tre.2019.06.004.

[6] Barreto, L., Amaral, A., Pereira, T. (2017). Industry 4.0 implications in logistics: An overview, *Procedia Manufacturing*, Vol. 13, 1245-1252, doi: 10.1016/j.promfg.2017.09.045.

[7] Lyu, Z., Lin, P., Guo, D., Huang, G.Q. (2020). Towards zero-warehousing smart manufacturing from zero-inventory just-in-time production, *Robotics and Computer-Integrated Manufacturing*, Vol. 64, Article No. 101932, doi: 10.1016/j.rcim.2020.101932.

[8] Urso, O., Chiacchio, F., Compagno, L., D'Urso, D. (2020). An RFID application for the process mapping automation, *Procedia Manufacturing*, Vol. 42, 8-15, doi: 10.1016/j.promfg.2020.02.017.

[9] Żywicki, K., Rewers, P. (2020). A simulation-based approach to study the influence of different production flows on manufacturing of customized products, *Advances in Production Engineering & Management*, Vol. 15, No. 4, 467-480, doi: 10.14743/apem2020.4.379.

[10] Sarı, T., Güleş, H.K., Yiğitol, B. (2020). Awareness and readiness of Industry 4.0: The case of Turkish manufacturing industry, *Advances in Production Engineering & Management,* Vol. 15, No. 1, 57-68, doi: 10.14743/apem2020.1.349.

[11] Focardi, R., Luccio, F.L., Wahsheh, H.A.M. (2019). Usable security for QR code, *Journal of Information Secuity and Applications*, Vol. 48, Article No. 102369, doi: 10.1016/j.jisa.2019.102369.

[12] Gjeldum, N., Mladineo, M., Crnjac, M., Veza, I., Aljinovic, A. (2018). Performance analysis of the RFID system for optimal design of the intelligent assembly line in the learning factory, *Procedia Manufacturing*, Vol. 23, 63-68, doi: 10.1016/j.promfg.2018.03.162.

[13] Arumugam, D.D., Engels, D.W. (2008). Impacts of RF radiation on the human body in a passive RFID environment, In: *Proceedings of 2008 IEEE Antennas and Propagation Society International Symposium*, San Diego USA, 1-4, doi: 10.1109/APS.2008.4619663.

[14] Lin, J.C. (2006). A new IEEE standard for safety levels with respect to human exposure to radio-frequency radiation, *IEEE Antennas and Propagation Magazine,* Vol. 48, No. 1, 157-159, doi: 10.1109/MAP.2006.1645601.

[15] TURCK Gmbh & Co. KG. BL ident® RFID System Planning and Engineering, from *https://www.turck.de/attachment/100000268.pdf*, accessed September 15, 2019.

[16] Qian, J.-P., Yang, X.-T., Wu, X.-M., Zhao, L., Fan, B.-L., Xing, B. (2012). A traceability system incorporating 2D barcode and RFID technology for wheat flour mills, *Computers and Electronics in Agriculture,* Vol. 89, 76-85, doi: 10.1016/j.compag.2012.08.004.

[17] Liang, K., Chen, X., He, R., Li, J., Okinda, C., Han, D., Shen, M. (2019). Development and parameter optimization of automatic separation and identification equipment for grain tracing systems based on grain tracers with QR codes, *Computers and Electronics in Agriculture,* Vol. 162, 709-718, doi: 10.1016/j.compag.2019.04.039.

[18] Liu, K., Bi, Y., Liu, D. (2020). Internet of things based acquisition system of industrial intelligent bar code for smart city applications, *Computer Communications,* Vol. 150, 325-333, doi: 10.1016/j.comcom.2019.11.044.

[19] Moreira, A., Silva, F.J.G., Correia, A.I., Pereira, T., Ferreira, L.P., de Almeida, F. (2018). Cost reduction and quality improvements in the printing industry, *Procedia Manufacturing*, Vol. 17, 623-630, doi: 10.1016/j.promfg.2018.10.107.

[20] Gaganis, C., Pasiouras, F., Voulgari, F. (2019). Culture, business environment and SMEs' profitability: Evidence from european countries, *Economic Modelling,* Vol. 78, 275-292, doi: 10.1016/j.econmod.2018.09.023.

[21] Vivanco Florido, J.S., Gonzalez Adame, M., Oropeza Tagle M.A. (2015). Financial strategies, the professional development of employers and performance of sme's (Aguascalientes case), *Procedia – Social and Behavioral Sciences,* Vol. 174, 768-775, doi. 10.1016/j.sbspro.2015.01.613.

[22] TimeStation. Turn any smartphone or tablet into a cloud-based time & attendance system, from *http://www.mytimestation.com*, accessed September 12, 2021.

[23] Cognex. Industrial cameras, from *https://www.cognex.com/*, accessed September 12, 2021.

[24] Vukicevic, A.M., Milic, V., Zabotti, A., Hocevar, A., De Lucia, O., Filippou, G., Frangi, A.F., Tzioufas, A., De Vita, S., Filipovic, N. (2019). Radiomics-based assessment of primary Sjögren's syndrome from salivary gland ultrasonography images, *IEEE Journal of Biomedical and Health Informatics,* Vol. 24, No. 3, 835-843. doi: 10.1109/JBHI.2019.2923773.

[25] Ho, C.-C., Wang, H.-P., Chiao, Y.-C. (2021). Deep learning based defect inspetion in TFT-LCD rib depth detection, *Measurement: Sensors*, Vol. 18, Article No. 100198, doi: 10.1016/j.measen.2021.100198.

[26] ISO/IEC 18004:2015, Information technology – Automatic identification and data capture techniques – QR code bar code symbology specification, from *https://www.iso.org/standard/62021.html*, accessed November 12, 2021.

[27] Kiurski, J.S., Marić, B.B., Oros, I.B., Kecić, V.S. (2017). The ecodesign practice in Serbian printing industry, *Journal of Cleaner Production*, Vol. 149, 1200-1209, doi: 10.1016/j.jclepro.2017.02.193.

[28] Visintin, F., (2012). Providing integrated solutions in the professional printing industry: The case of Océ, *Computers in Industry*, Vol. 63, No. 4, 379-388, doi: 10.1016/j.compind.2012.02.010.

[29] Vafaeenezhad, T., Tavakkoli-Moghaddam, R., Cheikhrouhou, N. (2019). Multi-objective mathematical modeling for sustainable supply chain management in the paper industry, *Computers & Industrial Engineering*, Vol. 135, 1092-1102, doi: 10.1016/j.cie.2019.05.027.

[30] Huin, S.F., Luong, L.H.S., Abhary, K. (2003). Knowledge-based tool for planning of enterprise resources in ASEAN SMEs, *Robotics and Computer-Integrated Manufacturing*, Vol. 19, No. 5, 409-414, doi: 10.1016/S0736-5845(02)00033-9.

[31] Erbe, H.-H. (2003). Manufacturing with low cost automation, *IFAC Proceedings Volumes,* Vol. 36, No. 7, 95-100, doi: 10.1016/S1474-6670(17)35813-5.

[32] Kopacek, P. (1992). 'Low-cost' factory automation, *IFAC Proceedings Volumes,* Vol. 25, No. 25, 27-33, doi: 10.1016/S1474-6670(17)49574-7.

[33] Github, Warehouse 4.0 solution for the pallet management in SMEs with QR-code, machine vision and IP cameras: A use-case from the printing industry, from *https://github.com/ArsoVukicevic/PalletManagement,* accessed September 12, 2021.

[34] Qian, J., Du, X., Zhang, B., Fan, B., Yang, X. (2017). Optimization of QR code readability in movement state using response surface methodology for implementing continuous chain traceability, *Computers and Electronics in Agriculture*, Vol. 139, 56-64, doi: 10.1016/j.compag.2017.05.009.

[35] Sengupta, J., Ruj, S., Das Bit, S. (2020). A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT, *Journal of Network and Computer Applications*, Vol. 149, Article No. 102481, doi: 10.1016/j.jnca.2019.102481.

APEM journal
Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 307–325
https://doi.org/10.14743/apem2021.3.402
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Increasing Sigma levels in productivity improvement and industrial sustainability with Six Sigma methods in manufacturing industry: A systematic literature review

Purba, H.H.[a], Nindiani, A.[b,*], Trimarjoko, A.[a], Jaqin, C.[a], Hasibuan, S.[a], Tampubolon, S.[a]

[a]Department of Industrial Engineering, Mercu Buana University, Jakarta, Indonesia
[b]Industrial Engineering Program, Buana Perjuangan University, Karawang, Indonesia

A B S T R A C T

Industrial sustainability is an important attribute and becomes a parameter of the business success. Quality improvement with an indicator of increasing process capability will affect productivity improvements and lead to industrial competitiveness and maintain industrial sustainability. The purpose of this paper is to obtain a relationship between the consistency of the DMAIC phase to increase the sigma level in productivity improvement and industrial sustainability. This paper applied for a systematic literature review from various sources of trusted articles from 2006 to 2019 using the keywords "Six Sigma, Productivity, and Industrial Sustainability." A matrix was developed to provide synthesis and summary of the literature. Six Sigma approach has been successful in reducing product variation, defects, cycle time, production costs, as well as increasing customer satisfaction, cost savings, profits, and competitiveness to maintain industrial sustainability. Extraction and synthesis in this study managed to obtain seven objectives value that found a consistent relationship between the DMAIC phase of increasing sigma levels, productivity, and industrial sustainability. The broad scope of Six Sigma literature is very beneficial for organizations to understand the critical variables and key success factors in Six Sigma implementation, which leads to substantial long-term continuous improvement, the value of money, and business.

A R T I C L E   I N F O

Keywords:
Manufacturing;
Sustainability;
Industrial sustainability;
Six Sigma;
Increase of Sigma level;
Productivity improvement;
Industrial competitiveness

**Corresponding author:*
aina.nindiani@gmail.com
(Nindiani, A.)

Article history:
Received 30 July 2020
Revised 24 October 2021
Accepted 26 October 2021

1. Introduction

Customer satisfaction in maintaining industrial sustainability is the key to business success in an increasingly competitive industrial era. Organizational / industrial competitiveness is one of the important attributes in increasing customer satisfaction to maintain industrial sustainability [1] (Ramana and Basavaraj. Maintaining industrial / company competitiveness can be performed by increasing productivity / sales, expanding marketing reach, and maintaining customer retention / loyalty by maintaining and improving the quality of its products [2, 3, 4]. Improving quality and productivity and maintaining customer satisfaction are important attributes in increasing competitiveness and maintaining industrial sustainability. Quality is an important issue in the highly competitive modern business world [1] (Ramana and Basavaraj, 2018); even quality is a key factor for consumers to decide on products and services offered by producers. Quality has customer satisfaction orientation [5] (Khawale *et al.*, 2017). In order to meet the quality of products expected by customers, an organization / company is always required to understand what is desired by customers by conducting various research and development to further design and create a product that has superior characteristics that are oriented to the desires and satisfaction of customers [6, 5]. Productivity is an indicator of success for an organization / company / country [7] (Maheshwari and Taparia, 2019). Productivity has the concept of how to produce or increase the production of goods and services optimally by utilizing resources efficiently [8] (Thorelli, 1960). Customer satisfaction is a response from the comparison of product performance with several standards before, during, and after consumption [9, 10, 11].

Customer satisfaction can be formed if the organization is able to provide product characteristics or attributes that meet customers' expectations and will have an impact on the reuse of products that have been used and can help improve the company's image through information about the products to other customers [12, 13, 14]. There are many strategies or methods or approaches that can be used in an effort to improve quality, productivity, and customer satisfaction, one of which is the Six Sigma approach. Six Sigma is a systematic and structured approach to increase performance / productivity and quality in meeting customer satisfaction to gain increased company profits [15, 16, 17, 18, 19].

Six Sigma has a systematic and structured method namely: Define, Measure, Analyze, Improve and Control (DMAIC) which is a stage of quality improvement with the concept of reducing the number of defects by up to 3.4 parts per one million opportunities which is very suitable in modern business that focuses on increasing customer satisfaction, productivity, and financial performance [15] (John and Areshankar, 2018). The basic concept of Six Sigma is to adhere to principles for process improvement through reducing variation, using statistical methods, focusing on customers, paying attention to processes, and management systems that focus on high yields that generate significant and continuous financial gains [20, 21, 22]. The level of readiness of the organization / company such as operating systems, measurement systems, employee involvement, environmental conditions, and the concept of continuous improvement greatly affect the success rate of Six Sigma implementation [23].

Six sigma success indicators in improving quality and productivity can be known by increasing the capability of the process (sigma levels) and the financial benefits obtained. Rahman *et al*. [17] in his research, succeeded in reducing defects such as broken stitches and open seam by 35 % and increasing sigma levels from 1.7 to 3.4. Rana and Kaushik [19] in the Six Sigma implementation have proven to reduce defects and increase productivity. In addition to other benefits that did not materialize (initiative, competitiveness), the DMAIC results showed that the defective washer thickness declined from 1550 PPM to near to 100 PPM within four months. In addition, Barbosa *et al*. [18] in his research using Six Sigma implementation (DMAIC), proved that it reduces defects and increase the Quality Rate of 41 % from 19 % to 60 % and Cp in Bead Apex process of > 1.33. Sardeshpande and Khairnar [24] used the Six Sigma method and succeeded in increasing the sigma level in the four wheels industry from 1.2 to 3.2. Subsequently, Ganguly [25] in his research, succeeded in reducing cycle time from 47 days to 20 days in one product cycle. Syafwiratama *et al*. [26] in his research, succeeded in increasing process capability (sigma levels) from 2.2 to 3.1 and making a profit of $ 18,394.2 per month. Furthermore, Malek and Desai [27] state that the Six Sigma approach is a paradigm that leads to business excellence based on the improvement of processes that have been proven to be widely adopted by various industries to respond to changes in customer desires or needs. This research succeeded in increasing sigma levels from 3.1 to 3.7 and reduced the reject rate from 15.50 % to 4.47 % or equivalent to 71.2 % and provide a cost-saving of INR 18,27,402.

Gandhi, Sachdeva, and Gupta [28] revealed that the Six Sigma approach is a systematic and scientific operations management methodology that aims to increase the ability of the production process by reducing waste. This research succeeded in reducing cylinder block defects from 28,111 to 9,708 DPMO and providing a profit of INR 12,56,640. Kosieradzka and Ciechańska [23] in their research with Six sigma implementation succeeded in increasing sigma levels from 3 sigma with DPMO = 33333.00 to 6 sigma with DPMO = 0 and Cp value = 1.91, Pp = 0.53, Cpk = 1.88, Ppk = 0.52 to Cp = 3.79, Pp = 3.32, Cpk = 3.34, Ppk = 2.29. Based on the statement of the research, there was a mistake of the DPMO value of Six Sigma should not be zero but 3.4. Chabukswar *et al*. [29] in his research stated that the basic concept of Six Sigma is to strengthen the process by modifying to increase the process capability to be able to produce defect-free products so that it can satisfy the customer. This research successfully identified the problem that was happening and managed to increase the process capability from 1.5 to 4 and reduce the process reworking 50 % and giving a profit of Rs 90 to 95 lakhs per year and increasing overall efficiency by 30 % in improving overall product quality. Morales *et al*. [30] in his research using Six Sigma, succeeded in reducing scrap from 18 % to 2 % in improving plant layout in an effort to increase productivity.

Chang and Wang [31] in their research using CPFR implementation (Collaborative planning, together with forecasting and replenishment) can increase forecast accuracy by 10-40 %, reduce inventory costs by 10-40 %, save transportation costs by 0.3 % and 1 % and increase customer service levels by 0.5-4 %. CPFR has been recognized as one of the most efficient methods to improve forecast accuracy, minimize inventory, increase service levels and reduce costs, research results using Six Sigma were known to reduce the Mean Absolute Percentage Error (MAPE) from 19.37 % to 5.26 %. They could reduce holdout products from 17.7 % to 5.18 % so that CPFR can help SCM to be better and can increase the confidence of business partners and increase the competitiveness of companies in an increasingly competitive era. Rahman and Talapatra [32] in their research using the Six Sigma approach in the casting industry, succeeded in reducing defect products (DPMO) from 609,302 to 304,651 and increasing the Sigma level from 1.2 to 2.0. In addition, Srinivasan *et al*. [33] used the Six Sigma method in the nozzle furnace industry and succeeded in increasing the sigma level from 3.31 to 3.67 and providing a cost saving of INR 0.125 million (the US $ 1953). Therefore, increasing the level of sigma proves that DMAIC is able to improve product quality, which results in cost reduction and increased competitiveness. According to the various literature studies above, it can be concluded that Six Sigma provides measurable indicators with the help of statistical methods and can be combined with other tools of analysis which are proven to be able to reduce product and process variability and be able to improve process capability through reducing defects, reducing process time so that it can reduce production costs, increase customer satisfaction and will certainly increase company profits in an effort to maintain the sustainability of the industry/organization on an ongoing basis.

2. Consistency of D-M-A-I-C phase, increase of Sigma level, productivity improvement and industrial sustainability

Six Sigma is a systemic, scientific, measurable, flexible and effective method of defining, measuring, analyzing, improving and controlling a problem to get better process capability by reducing process and product variations that aim to improve quality, increase productivity, increase satisfaction customers, increase competitiveness and also increase company profits in maintaining industrial sustainability. The structured method is known as the Define, Measure, Analyze, Improve, and Control (DMAIC) cycle, which has a concept of 3.4 defects per one million products [34, 35]. Broadly speaking, the intended DMAIC cycle can be understood in the description, as presented in Table 1. The table shows a general description of the DMAIC cycle performed in the problem-solving of each phase, including description and activities carried out in the implementation of the Six Sigma method.

Table 1 Description of the DMAIC cycle

Phase	Description of process activities
Define	Identify a problem/project *base on a business objective*
	Define project scope and goal bases on customer requirements
	Develop project charter and determined crucial to quality (CTQ)
Measure	Collect data, facts and carry out measurement systems
	Mapping process representations base on data collection
	Measure the process capability and study the differences (gaps) that occur

Table 1 Description of the DMAIC cycle (continuation)

Phase	Description of process activities
Analyze	Perform data analysis to find the cause of the problem Clarify the cause of the problem to find out whether the problem is a vital factor Determine the priority scale of each cause of the problem
Improve	Discussion to determine alternative improvements that can be implemented Carry out improvements according to the results of the discussion Verification of key variables in the implementation process
Control	Control process variations according to customer requirements Design monitoring and controlling strategies for improvement results Verify project objectives and plan for further improvement

2.1 Consistency of DMAIC phase

The Six Sigma approach is a systematic and scientific method of operations management that aims to improve the capability of the production process by reducing all wastes [28]. Six sigma approach with systematic phase namely DMAIC with statistical analysis tools has been proven to be widely adopted in various industries both manufacturing and services that can reduce various wastes, increase productivity, provide cost savings and increase company profits [36, 26, 33]. The consistency of the DMAIC phase in this study was analyzed from various studies to select research that was consistent in carrying out the entire DMAIC phase with standard tools that have been set on the Six Sigma method. The consistency of the implementation of the DMAIC phase using the Six Sigma method applied in various industries involving various research variables has succeeded in identifying, measuring, finding key factors, taking improvement actions and controlling problems so as to get better process capability (sigma levels) which is characterized by the reduction of defects, reduced cycle time, reduced accident rates, reduced breakdown / downtime, also improved planning accuracy with actual production and increased company / organizational profits. Through these indicators, the DMAIC Phase consistency proved to have a positive effect on improving quality and productivity and maintaining industrial sustainability.

2.2 Sigma levels

Process capability / sigma level is used to determine the performance capability of a process in producing goods or services based on established technical specifications [37, 38] so that it is known whether the process is within the expected limits or strict controls are needed for the ongoing process. Gupta *et al*. [39] in a study of the tire industry in India, revealed that Six Sigma is known to reduce the standard deviation from 2.17 to 1.69, and increase process capability (Cp) from 1.65 to 2.65 and Cpk from 0.95 to 2.66. This study shows that Six Sigma using DMAIC phase successfully lower the deviation standard from 2.17 to 1.69 which means that product variations can be suppressed so that product quality is better, the stability of the process can be improved as indicated by an increase in the Process capability index (Cp) from 1.65 to 2.65 and Cpk from 0.95 to 2.66. However, the result of the decreasing deviation standard value to 1.69 is quite high. It still gives the possibility to bring up unexpected conditions to happen. Thus, it can be concluded that an improved process capability index will stabilize the process and produce better quality so as to indicate better productivity.

Primanintyo *et al.* [40] in his research using Six Sigma and DOE as the Improvement method, succeeded in increasing the sigma levels in the Curing process of the tire industry in Indonesia from 3.092 sigma to 4.029 sigma. Gerger and Firuzan [22] in their research, explained that Six Sigma's main focus is to reduce the potential variability of processes and products by using a structured continuous improvement methodology, namely DMAIC. Six Sigma provides discipline, structure, and a foundation for solid decision making based on simple statistical analysis. The real strength of Six Sigma is simple because it combines the strength of people / management (project sponsor, project team leader) with the strength of the process (floor production / project team members) to get good process capability to improve organizational competitiveness. This research succeeded in reducing cycle time by 50 % from the previous conditions and can provide a profit of $ 167,895 per year. Zasadzień [41] revealed in the study that a complex and flexible system for achieving, maintaining, and maximizing business achievement is characterized by understanding customer needs and the use of facts, data, and statistical analysis in an organized manner and based on lean management and continually creating new ones. The better the solution by referring to the next process, aimed at minimizing the cost of poor quality while increasing customer satisfaction, this study succeeded in reducing downtime from 18 hours to 9 hours. According to these research examples, interpretation of sigma level improvement shows that process capability and level of process and product variation are better so that the potential for defects produced is more controlled. Customer satisfaction and company profitability can certainly be increased.

2.3 Productivity improvement

Productivity is an indicator of the success of an organization / company [7]. Productivity has the concept of how to produce or increase the production of goods and services optimally by utilizing resources efficiently [8]. An increase in productivity due to the literature can be interpreted as how companies / organizations utilize their resources in the form of tangible and intangible assets effectively and efficiently to obtain optimum profits. The Six Sigma method (DMAIC) is a structured method for identifying, analyzing cause and effect, as well as opportunities for improvement of an ongoing problem that aims to maintain the stability of the process to get product quality improvement and increase company profits [18]. The implementation of Six Sigma has been proven to reduce disability and increase productivity, in addition to other benefits that do not materialize, such as initiative and competitiveness [19]. Six Sigma is a scientific, systematic and superior method of responding to changes that occur in the business world and is able to improve quality and productivity through reducing the variety of processes and products [27, 28, 23]. There is a positive relationship between the Six Sigma approach to productivity improvement due to some literature, which is characterized by a decrease in variations in processes and defects that result in decreased cost production and increase cost-saving, competitiveness, and company profits.

2.4 Industrial sustainability

Industrial sustainability is the key to business success in an increasingly competitive industrial era, even in the last two decades, a total of 92 % of 200 companies published their industrial sustainability reports independently or in an integrated manner [42]. Sustainability reports have evolved over time and are considered an important component of organizational operations that are communicated annually to stakeholders through the sustainability report [43]. In this context, stakeholders increasingly demand transparency and accountability from companies regarding tangible sustainability performance [44]. This proves that the sustainability of an industry is an important attribute in the business world. Improving product / service quality and satisfaction, as well as maintaining customer loyalty, are strategies that can be carried out by the industrial world in maintaining the sustainability of the industry. [45, 46, 47] state that customer satisfaction affects the creation of customer loyalty, which will affect the company's revenue or profit and this is a major factor of industrial sustainability.

3. Materials and methods

The systematic literature review (SLR) method is a method of literature review that identifies, evaluates, and extracts / sites research findings that are useful in answering determining research questions [48]. Bolderston [49] revealed that a good literature review adopts several important rules, such as: (1) is able to extract new ideas from previous research by synthesizing and summarizing previous sources. New theories can be built from the evidence discussed and may provide new directions for future research. (2) A literature review may also facilitate the use of the best available evidence in daily practice to provide answers to research questions. This study is a systematic literature review on the implementation of Six Sigma and looks for the relationship between the consistency of the DMAIC phase to improve process capability (sigma levels) and productivity to increase customer satisfaction in maintaining industrial sustainability.

This study is initiated by collecting a variety of literature in the form of trusted articles from various sources, such as Google Scholar, Research Gate, Proquest, Academia.edu, and other sources using the keywords "Six Sigma, Productivity, and Industrial Sustainability." Furthermore, the works of literature obtained are then classified by name and country of author, year of publication, publisher, research variables, research objectives, tools used, and research results with the target of obtaining a literature review matrix that is useful for providing synthesis and summary of the literature that has been obtained to answer the research question as mentioned earlier. The phases used in this study can be seen in the Conceptual framework of literature review, as presented in Fig. 1.

Fig. 1 describes the whole phase and a review of each phase in this study that starts from defining the research goals, namely: The consistency of the DMAIC phase to increase process capability (sigma levels) and productivity improvement and industrial sustainability. The concept of sigma level, productivity, industrial sustainability, and Six Sigma approach are preliminary discussions followed by a collection of literature for the synthesis of the consistency of the DMAIC phase, sigma level, and productivity improvement, and industrial sustainability. Materials and Methods discuss the concept of Systematic Literature Review (SLR), followed by the conceptual framework of the literature review, which is an overview of the stages of this study. Stages of results and analysis presented include gaps / findings of the literature review, objective of the research, and recommendations for further research. In the last stage, the conclusions and limitations of the study and statements about this study are presented.

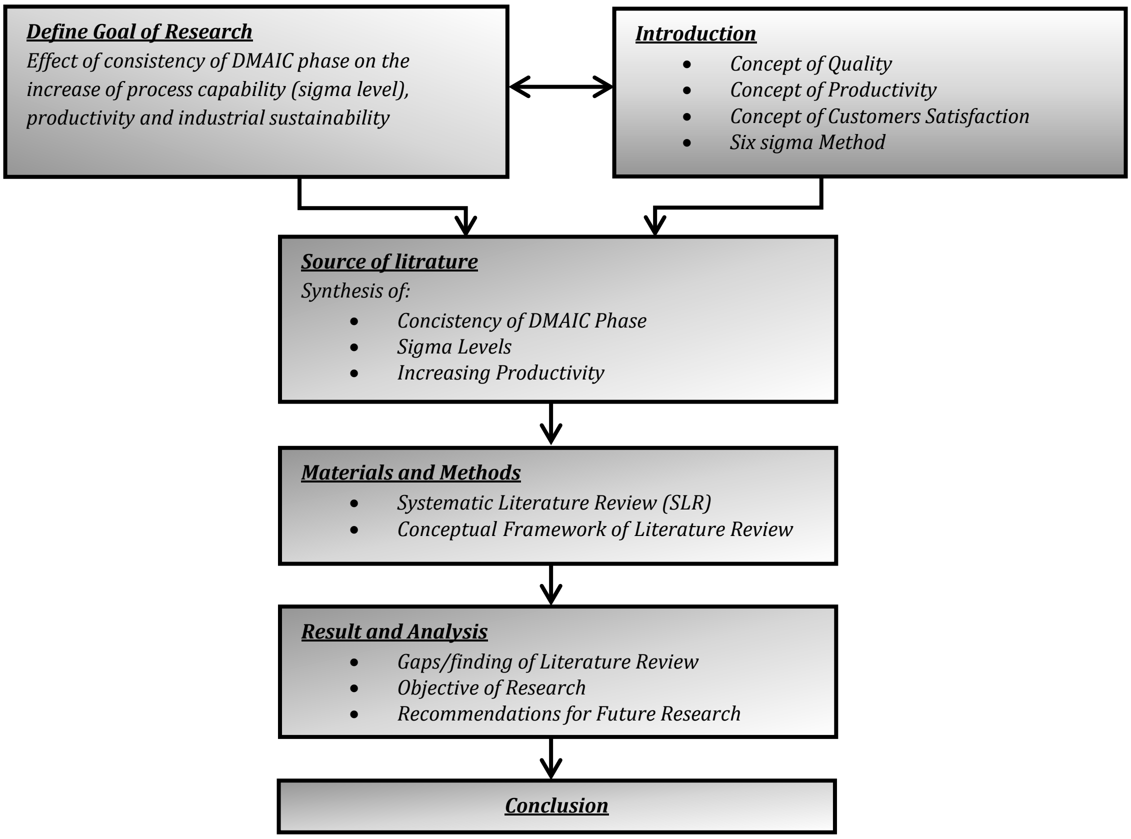


Fig. 1 Conceptual framework of the literature review

4. Results and discussion

4.1 Gaps/findings of literature review

The study is a literature review, which is a review of various journals and proceedings about the Six Sigma implementation in the manufacturing industry from various publishers starting from 2006 to 2019. Details of the literature and tools used in this study and the findings obtained can be seen in Table 2.

Table 2 Tools for literature review used and the research findings

No.	Author, Year, Country, Industry, Variable	Tools used					Result
		Define	Measure	Analyze	Improve	Control	
1	[1] Ramana and Basavaraj, 2018, India, Capacitor, Defects	VOC, VOB, Business mapping, SIPOC diagram, CTQ analyze	Brainstorming cause and effect diagram (CED)	Pareto diagram	Brainstorming purpose of solution (FMEA)	Standardization	Operational standard procedure, training, and control plan to reduce defects
2	[17] Rahman *et al.* 2018, Bangladesh, Garment, Defects	Voice of Customers (VOC)	Pareto diagram, sigma level measurement	Defect rate (DR), CED	DOE, ANOVA	Defect rate (DR)	Reduce defects by 35 %, increase sigma level 1.7 to 3.4
3	[19] Rana and Kaushik, 2018, India, Automotive SMEs, Defects	VOC, Flow diagram, SIPOC diagram	Gauge R&R study	RCA, Brainstorming Trial base on Comparison Worst of Worst (WOW) vs. Best of Best (BOB), Histogram	Corrective Actions with testing	PPM measurement, standardization	Reduce defects from 1550 PPM to almost 100 PPM in 4 months
4	[18] Barbosa *et al.* 2017, Italia, Tyre, Defects	Data collection of customer requirements (NC classification parameter)	CED, Brainstorming, Pareto diagram, $\bar{X}$ and R chart	Cp analysis	Taguchi DOE method	Validate the experimental results (cpk analysis and Quality Rate)	Increase quality rate from 19 % to 60 % and Cp > 1.33
5	[24] Sardeshpande and Khairnar, 2014, India, Automotive, Defects	SIPOC diagram, Process mapping	Control chart, Cost-benefit analysis, CTQ tree	Pareto Chart, Histogram, CED	Regeneration analysis, RCA, QFD	Process capability	Increase the sigma level from 1.2 to 3.2
6	[50] Kumar and Naidu, 2012, India, Garment, Employee Absenteeism	Data collection	Focused group discussion	Pareto diagram, CED	Corrective action	Control plan	Reduce employee absenteeism from 25-35 % to 11%
7	[25] Ganguly, 2012, India, Aluminium, Cycle Time	Historical data, VOC, SIPOC diagram	Defining all possible causes (FMEA), CTQ matrix, MSA	Scatter diagram, linear regression, ANOVA	DOE	Control plan (control chart, MSA)	Reduce cycle time of Rolling mill 47 days to 20 days
8	[26] Syafwiratama *et al.* 2016, Indonesia, Polyester, Defects	Data collection, Pareto diagram	Sigma level, four-block diagram	Brainstorming vital factor analysis (t-test)	DOE	X bar chart, Sigma level, four-block diagram	Increase sigma level from 2.2 to 3.1 and profit of $18,394.2 / month
9	[27] Malek and Desai, 2015, India, Casting, Defects	SIPOC, CTQ tree	Gauge R & R, p chart, capability process (sigma level)	CED, Regression analysis, why-why analysis	DOE, monitoring improvement p chart, COPQ analysis	Control plan, Operation-al standard procedure	Increase sigma level from 3.1 to 3.7 and profit of INR 18,27,402

Table 2 Tools for literature review used and the research findings (continuation)

No.	Author, Year, Country, Industry, Variable	Tools used					Result
		Define	Measure	Analyze	Improve	Control	
10	[51] Naidu, 2011, India, Steel, Breakdown Time	Brainstorming	Data collection of repair time and inspection time	Average data breakdown in one year, Pareto concept	Maintenance schedule	Regression & correlation analysis	Decrease breakdown from 92.42 hours To 59.0 hours
11	[28] Gandhi, Sachdeva, and Gupta, 2019, India, Casting, Reject of Product	CTQ analysis with VOC and Pareto diagram	Data collection, Pareto diagram, CED	P chart, t-test (Anderson-Darling), Why-why analysis	Three possible solutions	Sigma level measurement, overall rejection trend, t-test, Cost-Benefit Analysis	Reduce reject from 28,111 to 9,708 DPMO and provides a profit of INR 12,56,640
12	[23] Kosieradzka and Ciechańska, 2018, Poland, Various, Defects	SIPOCR	Data collection and control of element, Pareto concept	Capability analyzed Cp and Cpk and Pp and Ppk Cause Effect matrix, PFMEA	The 8D process	Cp and Cpk and Pp and Ppk measurement, SPC (control chart)	Increase sigma level from 3 to 6 and Cp = 1.91, Cpk = 1.88, to Cp = 3.79, Cpk = 3.34
13	[29] Chabukswar *et al.*, 2011, India, Pharmaceutical, Defects and Rework Time	VOC, Process flow diagrams, Baseline of the manufacturing process	Data collection	Verification of relationships and causality of factors	DOE	Control plan, SPC	Increase sigma level from 1.5 to 4.0, and provide a profit of Rs 90 to 95 lakh/year
14	[30] Morales *et al.* 2016, Mexico, Concrete Blocks (Construction), Breakdown, and Scrap	Process map, Pareto diagram	Measurements of the conveyor downtime	CED, CE matrix	Discussions relation of possible caused matrix	SPC	Reduce scrap from18 % to 2 %, breakdown molding from 87 to 43 cases
15	[52] Zhan, 2008, USA, Motor, Average Speed Variation	Establish a project scope with clear goals	CTQ tree, Normality test, speed motor measurement	DOE, Pareto chart	Modeling & Simulation	SPC	Reduce Deviation Standard by 74 %, target of 60 %
16	[53] Gajbhiye et al. 2016, India, Casting, Hazards and Risk Controls	SIPOC, SMART, PDCA, SWOT analysis and Cause and Effect analysis	5-S Audit (Safety, Sort, Straighten, Shine, Sustain)	5-Why + Safety Analysis, FMEA, RCA	5-S Audit, Safety Control Sheet, Safety Improvement Plan, and Post Kaizen EHS Checklist.	Safety Improvement Plan and Post Kaizen EHS Checklist	Decrease hazards and risk controls score machining section from 2.34 to 4.34
17	[31] Chang and. Wang, 2008, Taiwan, Various, Forecasting	Historical data, Mean absolute percentage error (MAPE)	Collect data sales	MAPE, autocorrelation function (ACF), partial autocorrelation function (PACF),control chart	ARIMA model	Control charts	Reduce MAPE from 19.37 % to 5.26 % and (holdout) from 17.7 % to 5.18 %
18	[54] Chang et al. 2012, Taiwan, Semiconductor, Production Planning	SIPOC, Questionnaire	t-test, Mann–Whitney test, and error measures	Sigma level, Control chart	Comparison planning vs. actual result	Cp measure, Questionnaire	Reduce average delay from 0.54 to 0.30

Table 2 Tools for literature review used and the research findings (continuation)

No.	Author, Year, Country, Industry, Variable	Tools used					Result
		Define	Measure	Analyze	Improve	Control	
19	[32] Rahman and Talapatra, 2015, Bangladesh, Casting, Defects	Data collection, Pareto diagram	Baseline performance (sigma level), Pareto diagram	CED	DOE, ANOVA comparison of data before vs. after improvement	Control plan	Reduce defects (DPMO) from 609,302 to 304,651, increase Sigma level from 1.2 to 2.0
20	[33] Srinivasan *et al.* 2016, India, Furnace, Defects	Pareto diagram, VOC, SIPOC diagram	Descriptive statistic, Gauge R and R	Brainstorming, CED	DOE, sigma level measurement	Comparison before after of data	Increase sigma level from 3.31 to 3.6 and provide cost-saving of INR 0.125 million (US$1,953)
21	[5] Khawale *et al.* 2017, India, Piston, Defects	SIPOC	VOC to CTQ	CED	DOE	Control plan	New standard operation procedure
22	[55] Purnama, Gunanto, and Sugengriadi, 2019, Indonesia, Manufacturing (other), Environment Management	Data collection	Pareto diagram	Stratify of data analysis, RCA	Corrective action (training)	SOP	Increase positive trend from 220 % to 700 %.
23	[56] El Hassani, Benlaajili and Nokra, 2017, Maroco, Sugar, Defects	5W 1H, Process map, a Black box of the process	Study of R & R, ANOVA, Control Charts, Process Capability	DOE, Pareto chart	Desirability study, The boxplot, Process Capability	Control Charts, standardization	Increase Average cp Aperture/ Opening Medium (OM) from 0.43 to 1.47 Coefficient of Variation (CV) from 0 to 1.5
24	[57] Soković, Pavletić and Krulčić, 2006, Slovenia, Automotive part, Cycle Time	Historical data	Pareto chart, discussion	FMEA, ANOVA, Correlation matrix	Brainstorming, Experiment, Cp analysis, and gage R & R	Control plan	Reduce production time, control time ($ 72,000)
25	[20] Hassan, 2013, Egypt, Wire, Waste Reduction	SIPOC Diagram	Process Mapping, Data Collection (Pareto chart), Sigma level calculations, Down Time Measurements	CED, AHP	Action plan	Control plan	Increase the sigma level from 3.2 to 3.6
26	[58] Rathilall and Singh, 2018, South Africa, Automotive Part, Key Factor LSS	Questionnaire	validity and reliability measure with Cronbach's alpha test	Gap analysis	Pearson's Chi-squared test	Gap analysis	Found six-item Critical factors of LSS implementation
27	[41] Zasadzień, 2017, Poland, Plastic, Downtime	Data collection	Process map	FMEA	Creation matrix	Control plan, standardization	Reduce downtime from 18 hours to 9 hours
28	[15] John and Areshankar, 2018, India, Machining, Defects	Brainstorming, Pareto diagram	Normality test, Cp analysis	Individual chart, CED	DOE, ANOVA	Individual chart, Cp analysis, Pareto diagram	Increase Cp of diameter and thickness from 0.27 and 0.35 to 1.03 and 1.69

Table 2 Tools for literature review used and the research findings (continuation)

No.	Author, Year, Country, Industry, Variable	Tools used					Result
		Define	Measure	Analyze	Improve	Control	
29	[16] Anand *et al.* 2007, India, Automotive Part, Reject of Product	FGD, data history	Critical to Quality (CTQ), QFD	DOE, Control Chart	Fuzzy-rule, Anova	CPk analysis, control chart	Increase punch load Cpk from 0.447 to 1.33
30	[59] Kaushik and Khanduja, 2007, India, Thermal Power, Defects	SIPOC diagram	Gauge R&R	Run chart, process capability analysis, CED, Bar chart	SWAS (Steam water analysis system), Training, action plan	Documentation	Increase sigma level 2.0 to 3.0 and reduce COPQ Rs from 304,77 lakh to Rs 331,2 lakh per year
31	[60] Desai and Prajapati, 2017, India, Plastic, Defects	SIPOC diagram, Pareto diagram, process map	Data collection	Brainstorming, CED, Multi-voting, Cause validation, Why-why analysis	Preventive maintenance, SOP, kaizen	SOP with visual aids	Defects reduced and gave savings of INR 10.80 lacs
32	[22] Gerger and Firuzan, 2016, Turkey, Aerospace, Cycle Time	Data collection	Production flow chart, CED	Effect of weight on human bone	Comprising of ex and new apparatus weight	Control plan	Reduce cycle time by 50 % and provide a profit of $ 167,895 / year.
33	[21] Sławik et al. 2010, Poland, Automotive Part, The Variation Rate	Modeling	Comparison differences between time scales in which aeration processes occur	CED	DOE analysis, Simulation	Gage R&R	Reduce variation of absorber lake from 92 % to 2.4 % with a variation of tolerance base Gage R & R 710 by 68 %
34	[36] Hussain, Jamshaid and Sohail, 2014, Pakistan, Textile, Defects	Data collection, flow process diagram	Cp measurement	CED, Pareto chart	Risk matrix table, corrective actions	SPC, FMEA	Increase sigma level from 2.2 to 3.0 with a profit of $26,000 per month
35	[61] Gijo et al., 2011, India, Automotive, Defects	SIPOC diagram, process map	Gage R & R, Kappa statistic	CED, cause validation plan, Pareto, cause validation plan, process capability analysis	DOE, Taguchi S/N Ratio, ANOVA	Cause solution matrix, standardization	Defects reduction from 16.6 to 1.19 %, sigma level improvement from 2.47 to 3.76, annualized savings about US$2.4 million
36	[62] Anderson and Kovach, 2014, USA, Construction, Repair rate	SIPOC diagram, Process map	Histogram, Pareto charts	CED, FMEA	Prioritization matrices, training,	Control plan, run chart, visual inspection, checklist	The weld repair rate decreased by more than 25 %, which translated into a savings of $90,000
37	[63] Sharma et al, 2018, India, Amplifier, Defects	SIPOC diagram	P-chart, Pareto diagram	CED, Current Reality Tree (CRT) map	Training, process instruction guide	P-chart, Pareto diagram	Improving the sigma level of the anodising process from 3.62 to 3.91

Table 2 Tools for literature review used and the research findings (continuation)

No.	Author, Year, Country, Industry, Variable	Tools used					Result
		Define	Measure	Analyze	Improve	Control	
38	[64] Kumaradivel and Natarajan, 2013, India, Flywheel Casting, Defects	SIPOC diagram	CTQ tree, Cause-and-effect matrix, Pareto chart	Cp analysis, FMEA, Key Process Input and Output Variables (KPIV, KPOV), Pareto diagram	Experimental design matrix, ANOVA	Histogram, SPC, PDCA	The rejection percentage declined to 4.69 from 6.94
39	[65] Lo *et al.*, 20019, Taiwan, Optic Elements, Defects	Abberation measurement for good-precision molded lenses, process analysis	Process capability	CED, Taguchi, ANOVA	Optimal combination of process capabilities	Standardization, built-in monitoring system	C_{PU} for optical lenses is enhanced from 0.57 to 1.75
40	[66] Chen *et al.*, 2009, USA, Plasma Cutting, Defects	Brainstorming	Cause and Effects Matrix, System capability measurements	Multi-Vari Analysis, T-tests	Taguchi	Control Plan, Continuous improvement	The optimal setting combination gave no defects from the 30 plasma-cut holes in the confirmation run. It maintains feed rate for productivity and improves quality.

Based on Table 2, the sources of literature above can be grouped as presented in Fig. 2a to Fig. 2e:

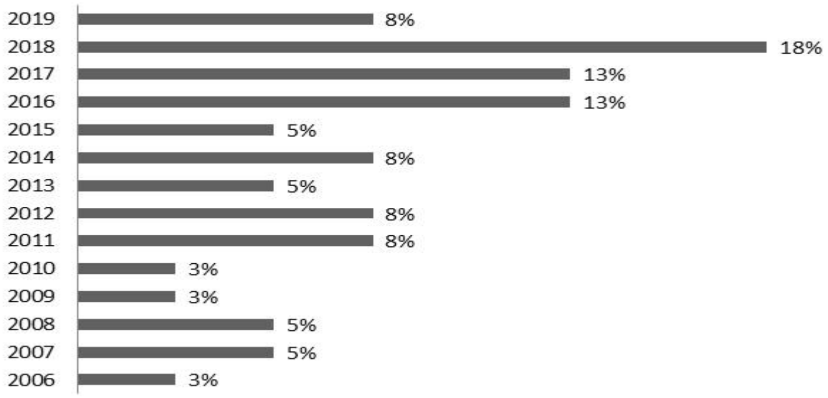


Fig. 2a Grouped sources of literature: Year of Published

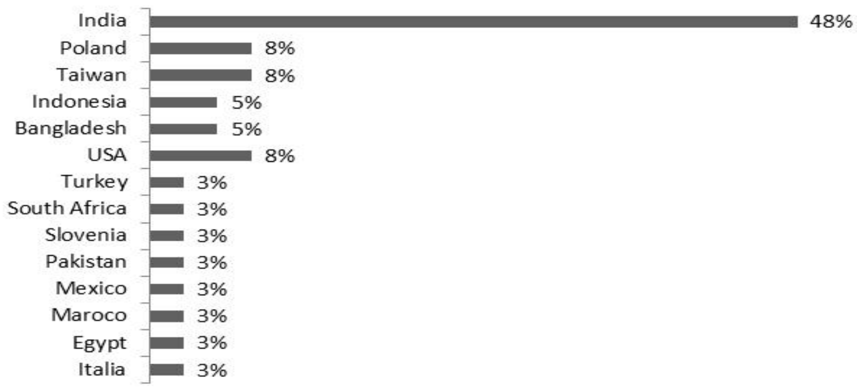


Fig. 2b Grouped sources of literature: Countries of Author

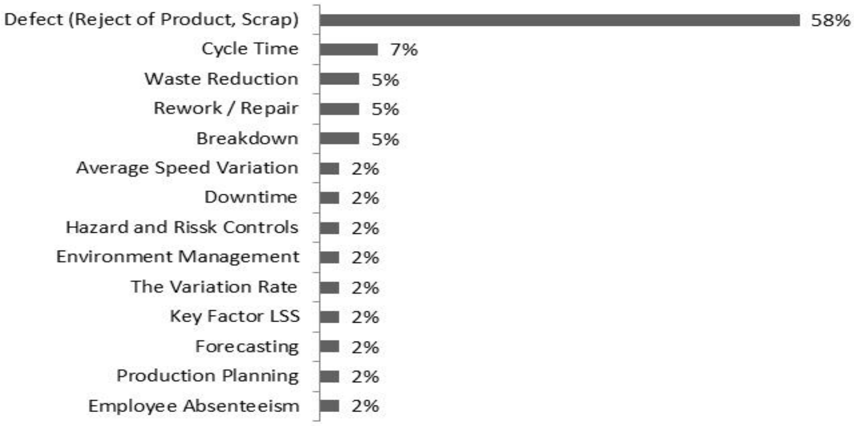


Fig. 2c Grouped sources of literature: Research Variable

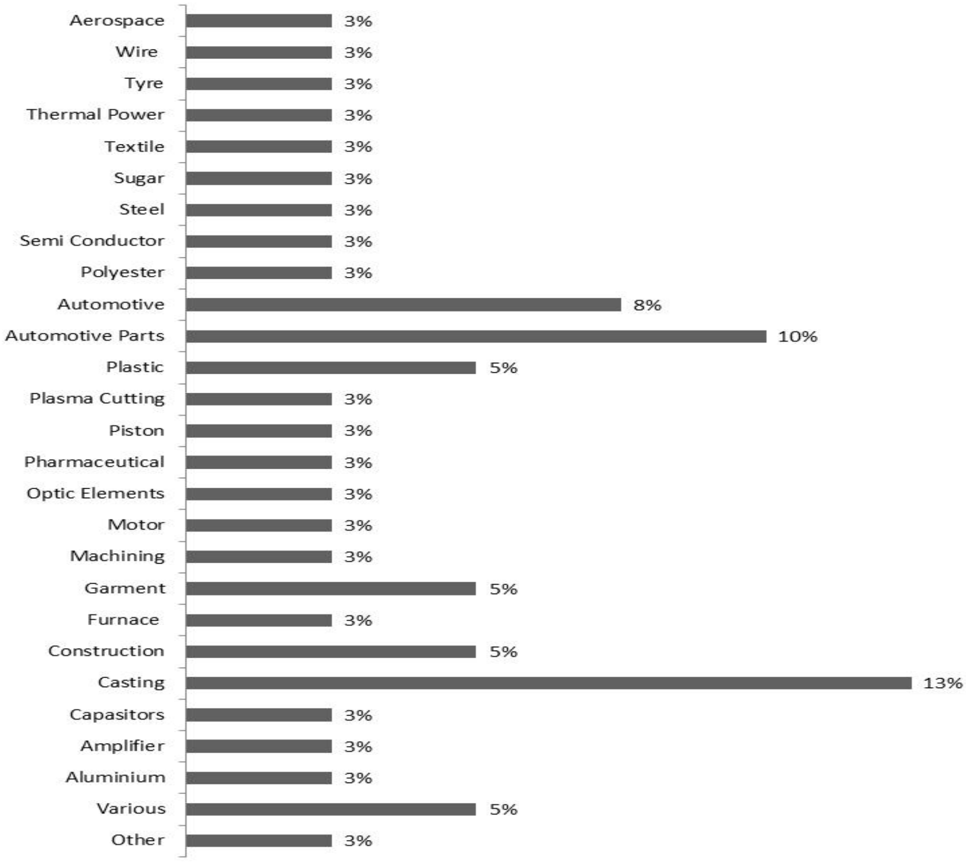


Fig. 2d Grouped sources of literature: Industries Type

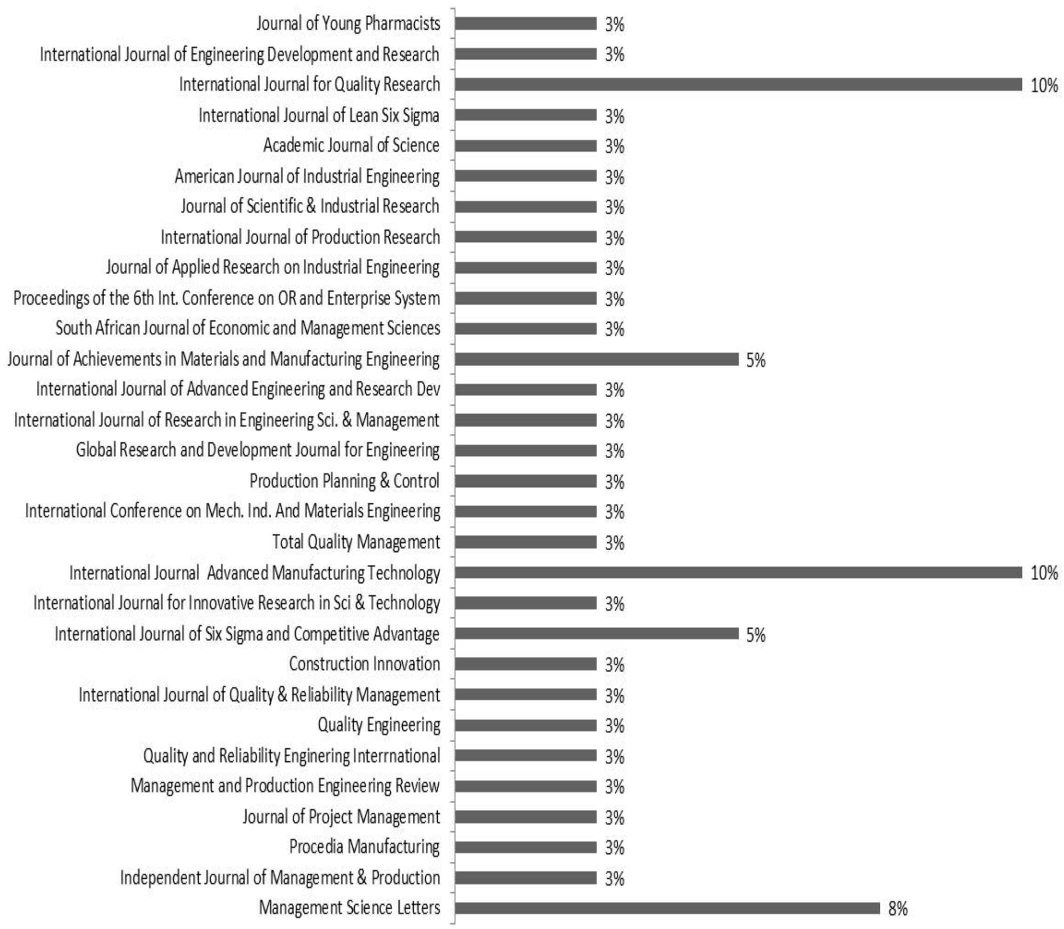


Fig. 2e Grouped sources of literature: Publisher of Research

Table 2 and Fig. 2a to Fig. 2e discuss Six Sigma as review literature starting from 2006 to 2019 from 14 countries with 27 types of industries and involved 14 research variables and 30 publishers. The implementation of Six Sigma based on existing literature studies shows varied results. This is thought to have occurred due to differences in researchers' analytical skills and organizational / company readiness in implementing Six Sigma. The company's readiness includes; operating systems, measurement systems, information integration systems, employee involvement, the concept of continuous improvement used, support for environmental conditions, and management commitment. Based on the literature, it is expected to represent the implementation conditions and the results obtained in the implementation of Six Sigma concept. It will be able to answer the objectives of this study by proving that the consistency of the DMAIC phase in Six Sigma will be able to increase the productivity of an organization / company. The Six Sigma literature is presented to illustrate the broad scope of the application of the concept. It is very beneficial for organizations to understand the critical variables and key success factors in the implementation of Six Sigma programs, which leads to substantial long-term continuous improvement for performance, the value of money, and business. This paper presents the facts of the successful implementation of Six Sigma, which is proven by data and presents the key success factors, variables, and interrelationships. Fig. 3 describes the research results and variables.

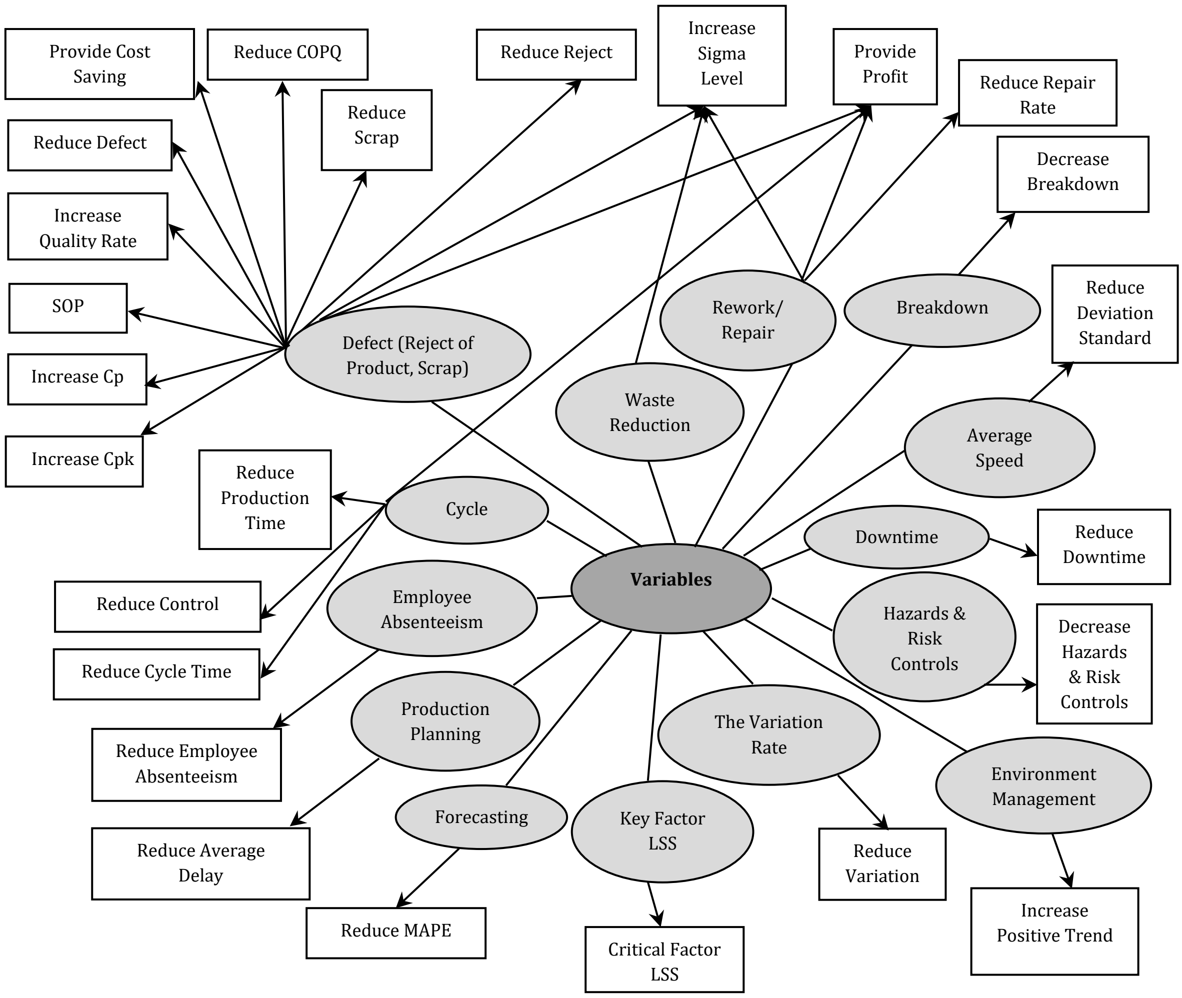


Fig. 3 The research result and variables

4.2 Objective value of the study

Systematic DMAIC phase

According to the literature used in this study, the Six Sigma phase can be synthesized with the tools used and the expected goals in implementing the DMAIC phase, as presented in Table 3.

Table 3 Systematic DMAIC phase

Phase	Description activity	Tools	Main of Goals
Define	• Define the problem • Determine CTQ • Determine improvement targets • Create a charter project	VOC/VOB, SIPOC diagram/ process map, Historical data, Pareto diagram, Questionnaire, Brainstorming.	• Obtain the CTQ • Build a team of improvement • Obtain improvement targets
Measure	• Collect data and facts • Mapping to represent data • Measurement of the condition base on data and facts obtained	Gage R &R, Pareto diagram, Control chart, Cp (sigma level) measurement, four-block diagram, descriptive dan non-descriptive statistic analysis.	• Obtain a process capability (sigma level) that represents the condition before improvement.
Analyze	• Perform analysis base on data and fact • Vital factor testing • Planning improvements	Brainstorming, CED, CE Metrik, RCA analysis, Comparison WOW vs. BOB, Scatter diagram, linear regression, Anova, why-why analysis, three possible analysis, FMEA, gaps analysis.	• Know the potential causes of the problem • Know the main causes (vital factors) of the problem • Develop an effective improvement plan.

Table 3 Systematic DMAIC phase (continuation)

Phase	Description activity	Tools	Main of Goals
Improve	• Discussion to determine alternative improvements that can be implemented • Carry out improvements according to the results of the discussion • Verification of key variables in the implementation process	DOE, FMEA, Gage R &R, Pareto diagram, Control chart, Cp (sigma level) measurement, four-block diagram, Descriptive and non-Descriptive statistic analysis, Three possible solutions, Corrective action.	• Carry out improvement • Strive for the effectiveness of improvements by looking at the comparison of the results of improvements with conditions before improvement
Control	• Control process variations according to customers' requirements • Design monitoring and controlling strategies for improvement results • Verification of objective and standardized projects, sharing new standards and determining the further project	SPC/Control chart, Control plan, SOP, Historical data, Questionnaire, Brainstorming.	• Obtain a controlled process. • Obtain new standards / documentation from the improvement process • Ensuring new standards are known and implemented throughout the entire organization. • Make further improvement plans.

Increase of Sigma Levels, Productivity Improvement and Industrial Sustainability Using the Six Sigma Method

The consistency of the DMAIC phase in Six Sigma implementation has been proven and has succeeded in increasing quality, reducing unnecessary production costs, and increasing productivity. Khawale *et al.* [5] in his study, stated that the DMAIC (Six Sigma Methodology) approach could be used to reduce defects and increase productivity. Six Sigma is a method that results in business excellence with a focus on the needs and expectations of customers. It is the key to the success of this method, based on facts and analysis with measurable statistical methods so that the results can be accounted for in managing businesses currently both manufacturing and services. The Six Sigma implementation is directly related to the company's finances, resulting in customer satisfaction being the target of this method and with innovative ways to exceed the expectations / desires and satisfaction of the customers. Jacob and Jenson [67] in their study of the tire industry in India using the VSM and Six Sigma methods, succeeded in increasing speed calendering machines and reducing cycle time from 17 hours 37 minutes to 16 hours 15 minutes. This study explains that by running the entire DMAIC phase combined with the Value Stream Mapping (VSM) method, it can reduce the cycle time of the calendering process, which means that the productivity of the calendering process can be improved.

Soković, Pavletić, and Krulčić [57] in their research on the Automotive industry explained that Six Sigma with the help of tools of analysis is proven to be able to reduce product and process variability and be able to improve process capability through reducing defects and reducing cycle times so as to reduce production costs and certainly will increase company profits. In general, increases obtained through reduced production time and control time can provide an annual profit of $ 72,000. The expected annual profit from implementing this system is $ 100.

Based on several research findings using the Six Sigma method as explained earlier, it is proved that the consistency of the implementation of the structured phases in Six Sigma, namely DMAIC phase, may provide positive results in solving problems. It may improve process capability / sigma levels and productivity as indicated by decreasing variations, defects, cycle time, customer complaints, non-value-added, and production cost, as well as increasing product quality, customer satisfaction, cost-saving / profit, competitiveness and maintaining industrial sustainability. Fig. 4 illustrates the relationship between the consistency of the DMAIC phase with increasing sigma levels, productivity, and industrial sustainability.

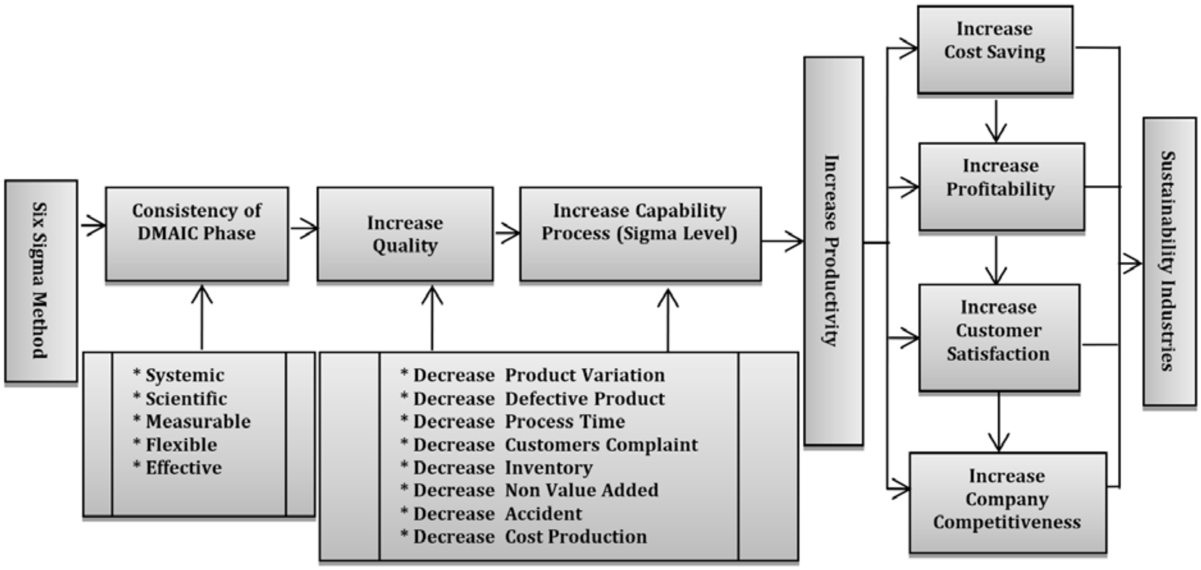


Fig. 4 Relationship/effects of Six Sigma method on increasing Sigma levels, productivity and industrial sustainability

According to the literature review on the consistency of DMAIC in the implementation of Six Sigma based on the literature that has been obtained and after going through analysis and synthesis based on the rules of systematic literature review (SLR), it can be concluded that the objective values obtained from this study are as follows:

1. Process capability (sigma levels) is an indicator of process stability or capability to produce a quality product. The higher the value of process capability (sigma levels), the process will be able to produce products with better quality, vice versa.
2. Productivity is an indicator of the success of an organization / company that has the concept of how to produce or increase the production of goods and services optimally by utilizing resources in the form of tangible and intangible assets effectively and efficiently.
3. Implementation of Six Sigma in a business organization system is a systemic approach (has definite stages), *Scientific* (based on data and facts), *Measurable* (has definite measurement standards with statistical methods), *Flexible* (can be combined with methods and other tools of quality) and *Effective* (is able to increase productivity at a low cost by reducing defects) to revolutionize the scope and use of quality systems in the business currently.
4. Six Sigma is a complex and flexible method / system for achieving, maintaining, and maximizing business achievement that is characterized by understanding customer needs by using facts, data, and statistical analysis and is based on organized management to perform continuous improvement.
5. Six Sigma implementation has varying results depending on the level of readiness of the organization/company, operating system, measurement system, information integration system, employee involvement, the concept of continuous improvement, environmental support, and top management commitment.
6. There are two main benefits of implementing Six Sigma to the effectiveness of the company. (1) Direct benefits: these benefits are in the form of financial side obtained from the Six Sigma implementation which is characterized by increasing quality and productivity which will definitely provide cost savings and increase the profit of the company / organization; (2) Indirect benefits: these benefits are in the form of the non-financial side including increased teamwork, increased sense of belonging among employees, increased employee competence, increased employee initiative, increased quality, increased trust in business relationships, which will further increase competitiveness in maintaining the continuity of the company / organization.
7. Quality, process capability (sigma levels), and productivity are attributes of customer satisfaction to maintain competitiveness and industrial sustainability and have a positive relationship and are directly proportional between these attributes.

Recommendation

Referring to the literature that has been reviewed, various previous studies show that the implementation of Six Sigma is more likely to provide tangible / direct benefits such as; reduce defects, reduce downtime, reduce cycle time, forecast accuracy and others that will all provide financial benefits, yet very rarely research that discusses the hidden/indirect benefits of Six Sigma that will provide non-financial benefits. It is recommended that more subsequent studies discuss research methodology that will provide results in terms of hidden / indirect benefits so that research on Six Sigma is more varied.

5. Conclusion

This paper is a systematic literature review on the consistency of DMAIC using the Six Sigma method from various previous studies in the manufacturing industry of 2006 to 2019 from 14 countries with 27 types of industries and involving 14 research variables and 30 publishers. According to the literature reviews, it shows different results from each research that indicate differences in the level of analysis ability of researchers and the level of readiness of the company / organization in implementing Six Sigma, however, overall the Six Sigma approach has been successful in reducing product variation, reducing defects, reducing cycle time, reducing production costs and increasing customer satisfaction, providing cost savings, increasing profits and increasing competitiveness in order to maintain the sustainability of the company / industry. The study also succeeded in obtaining seven objective values , which are the main results of this study and managed to find a consistent relationship between the DMAIC phase of increasing sigma levels, increasing productivity, and industrial sustainability as the research questions in this study.

References

[1] Raman, R.S., Basavaraj, Y. (2019). Defect reduction in a capacitor manufacturing process through six sigma concept: A case study, *Management Science Letters*, Vol. 9, 253-260, doi: 10.5267/j.msl.2018.11.014.

[2] Tsoukatos, E., Rand, G.K. (2006). Path analysis of perceived service quality, satisfaction and loyalty in Greek insurance, *Managing Service Quality: An International Journal*, Vol. 16, No. 5, 501-519, doi: 10.1108/09604520 610686746.

[3] John, J. (2011). An analysis on the customer loyalty in telecom sector: Special reference to Bharath Sanchar Nigam limited, India, *African Journal of Marketing Management,* Vol. 3, No. 1, 1-5.

[4] Loke, S.-P., Taiwo, A.A., Salim, H.M., Downe, A.G. (2011), Service quality and customer satisfaction in a telecommunication service provider, In: *Proceedings of 2011 International Conference on Financial Management and Economics*, Singapore, 23-29.

[5] Khawale, Y.S., Jagdale, S.C., Jambhalekar, S.S., Kale, S.S., Patil, V.H. (2017). Implementation of six sigma methodology for piston rod manufacturing, *Global Research and Development Journal for Engineering*, Vol. 2, No. 3, 14-18.

[6] Hadi, H.A., Purba, H.H., Indarto, K.S., Simarmata, R.G.P., Putra, G.P., Ghazali, G., Aisyah, S. (2017) The implementation of quality function deployment (QFD) in tire industry, *ComTech*, Vol. 8, No. 4, 223-228, doi: 10.21512/comtech.v8i4.3792.

[7] Maheshwari, M., Taparia, P. (2019). Measurement of material productivity: A case study of pharmaceutical sector companies included in Nifty 50, *Productivity*, Vol. 60, No. 2, 175-194, doi: 10.32381/PROD.2019.60.02.7.

[8] Thorelli, H.B. (1960). The tantalizing concept of productivity, *American Behavioral Scientist,* Vol. 4, No. 3, 6-11, doi: 10.1177/000276426000400303.

[9] Levesque, T., McDougall, G.H.G. (1996). Determinants of customer satisfaction retail banking, *International Journal of Bank Marketing*, Vol. 14, No. 7, 12-20, doi: 10.1108/02652329610151340.

[10] Srivastava, S., Bhatnagar, A. (2013). Impact of customer care services on customer satisfaction – A study of mobile phone subscribers of u.p. (east) circle, *International Journal of Management Research and Review*, Vol. 3, No. 1, 2224-2242.

[11] Otim, S., Grover, V. (2006). An empirical study on web-based services and customer loyalty, *European Journal of Information Systems,* Vol. 15, No. 6, 527-541, doi: 10.1057/palgrave.ejis.3000652.

[12] Chang, H.H., Wang, H.-W. (2011). The moderating effect of customer perceived value on online shopping behaviour, *Online Information Review*, Vol. 35, No. 3, 333-359, doi: 10.1108/14684521111151414.

[13] Parasuraman, A., Grewal, D. (2000). The impact of technology on the quality-value-loyalty chain: A research agenda, *Journal of the Academy of Marketing Science*, Vol. 28, No. 1, 168-174, doi: 10.1177/0092070300281015.

[14] Kheng, L.L., Mahamad, O., Ramayah, T., Mosahab, R. (2010). The impact of service quality on customer loyalty: A study of banks in Penang, Malaysia, *International Journal of Marketing Studies*, Vol. 2, No. 2, 57-66, doi: 10.5539/ijms.v2n2p57.

[15] John, B., Areshankar, A. (2018). Reduction of rework in bearing end plate using six sigma methodology: A case study, *Journal of Applied Research on Industrial Engineering,* Vol. 5, No. 1, 10-26, doi: 10.22105/jarie.2018.120059.1033.
[16] Anand, R.B., Shukla, S.K., Ghorpade, A., Tiwari, M.K., Shankar, R. (2007). Six sigma-based approach to optimize deep-drawing operation variables, *International Journal of Production Research*, Vol. 45, No. 10, 2365-2385, doi: 10.1080/00207540600702308.
[17] Rahman, A., Shaju, S.U.C., Sarkar, S.K., Hashem, M.Z., Hasan, S.M.K., Islam, U. (2018). Application of six sigma using define measure analyze improve control (DMAIC) methodology in garment sector, *Independent Journal Of Management & Production*, Vol. 9, No. 3, 810-826, doi: 10.14807/ijmp.v9i3.732.
[18] Barbosa, B., Pereira, M.T., Silva, F.J.G., Campilho, R.D.S.G. (2017). Solving quality problems in tyre production preparation process: A practical approach, *Procedia Manufacturing*, Vol. 11, 1239-1246, doi: 10.1016/j.promfg.2017.07.250.
[19] Rana, P., Kaushik, P. (2018). Initiatives of six-sigma in an automotive ancillary unit: A case study, *Management Science Letters,* Vol. 8, 569-580, doi: 10.5267/j.msl.2018.4.033.
[20] Hassan, M.K. (2013). Applying lean six sigma for waste reduction in a manufacturing environment, *American Journal of Industrial Engineering*, Vol. 1, No. 2, 28-35.
[21] Sławik, D., Czop, P., Król, A., Wszołek, G. (2010). Optimization of hydraulic dampers with the use of design for six sigma methodology, *Journal of Achievements in Materials and Manufacturing Engineering,* Vol. 43, No. 2, 676-683.
[22] Gerger, A., Firuzan, A.R. (2016). Use of six sigma method as a process improvement technique: A case study in aerospace industry, *Academic Journal of Science*, Vol. 6, No. 1, 191-202.
[23] Kosieradzka, A., Ciechańska, O. (2018). Impact of enterprise maturity on the implementation of six sigma concept, *Management and Production Engineering Review*, Vol. 9, No. 3, 59-70, doi: 10.24425/119535.
[24] Sardeshpande, A.K., Khairnar, H.P. (2014). Application of six sigma in four wheeler platform truck with quality tools and techniques, *International Journal of Engineering Development and Research*, Vol. 2, No. 1, 126-134.
[25] Ganguly, K. (2012). Improvement process for rolling mill through the DMAIC six sigma approach, *International Journal for Quality Research*, Vol. 6, No. 3, 221-231.
[26] Syafwiratama, O., Hamsal, H., Purba, H.H. (2017). Reducing the nonconforming products by using the six sigma method: A case study of a polyester short cut fiber manufacturing in Indonesia, *Management Science Letters,* Vol. 7, 153-162, doi: 10.5267/j.msl.2016.12.001.
[27] Malek, J., Desai, D. (2015). Reducing rejection/rework in pressure die casting process by application of DMAIC methodology of six sigma, *International Journal for Quality Research,* Vol. 9, No. 4, 577-604.
[28] Gandhi, S.K., Sachdeva, A., Gupta, A. (2019). Reduction of rejection of cylinder blocks in a casting unit: A six sigma DMAIC perspective, *Journal of Project Management*, Vol. 4, 81-96, doi: 10.5267/j.jpm.2019.1.002.
[29] Chabukswar, A.R., Jagdale, S.C., Kuchekar, B.S., Kothawade, H.S., Kuckekar, A.B., Joshi, V.D., Deshmukh, G.R., Lokhande, P.D. (2011). Six sigma: Process of understanding the control and capability of ranitidine hydrochloride tablet, *Journal of Young Pharmacists,* Vol. 3, No. 1, 15-25, doi: 10.4103/0975-1483.76415.
[30] Noriega Morales, S., Valles Ch., A., Torres-Argüelles, V., Martínez, G.E. Hernández, G.A. (2016). Six sigma improvement project in a concrete block plant, *Construction Innovation*, Vol. 16, No. 4, 526-544, doi: 10.1108/CI-01-2015-0003.
[31] Chang, K.K., Wang, F.K. (2008). Applying six sigma methodology to collaborative forecasting, *The International Journal of Advanced Manufacturing Technology,* Vol. 39, 1033-1044, doi: 10.1007/s00170-007-1276-5.
[32] Rahman, M.A., Talapatra, S. (2015). Defects reduction in casting process by applying six sigma principles and DMAIC problem solving methodology (A case study), In: *Proceedings of International Conference on Mechanical, Industrial and Materials Engineering 2015 (ICMIME2015)*, Rajshahi, Bangladesh, 1-6.
[33] Srinivasan, K., Muthu, S., Devadasan, S.R., Sugumaran, C. (2016). Enhancement of sigma level in the manufacturing of furnace nozzle through DMAIC approach of six sigma: A case study, *Production Planning & Control,* Vol. 27, No. 10, 810-822, doi: 10.1080/09537287.2016.1143130.
[34] Adeyemi, Y., Needy, K.L. (2006). An analysis of six sigma at small vs. large manufacturing companies, In: *Proceedings of IIE Annual Conference,* Norcross, USA, 1-6.
[35] Pyzdek, T. (2003). *The six sigma handbook*, Mc Graw Hill, New York, USA.
[36] Hussain, T., Jamshaid, H., Sohail, A. (2014). Reducing defects in textile weaving by applying six sigma methodology: A case study, *International Journal of Six Sigma and Competitive Advantage*, Vol. 8, No. 2, 95-104, doi: 10.1504/IJSSCA.2014.064254.
[37] Purba, H.H., Aisyah, S. (2017). *Quality improvement and lean six sigma: Meningkatkan kulitas produk dan kinerja perusahaan menuju zero defect,* Expert, Yogyakarta, Indonesia.
[38] Ryall, J., Krunithof, J. (2001). *The quality systems handbook*, Consesus Books, Sydney, Australia.
[39] Gupta, V., Jain, R., Meena, M.L., Dangayach, G.S. (2018). Six-sigma application in tire-manufacturing company: A case study, *Journal of Industrial Engineering International*, Vol. 14, No. 3, 511-520, doi: 10.1007/s40092-017-0234-6.
[40] Primanintyo, B., Syafei, Y.M., Luviyanti, D. (2016). Analisis penurunan jumlah defect dalam proses tire-curing dengan penerapan konsep six sigma, *Journal of Industrial Engineering*, Vol. 1, No. 2, Bekasi, Indonesia.
[41] Zasadzień, M. (2017). Application of the six sigma method for improving maintenance processes – Case study, In: *Proceedings of the 6th International Conference on Operations Research and Enterprise Systems (ICORES 2017),* Porto, Portugal, 314-320, doi: 10.5220/0006184703140320.
[42] KPMG. (2015). The KPMG survey of corporate responsibility reporting 2015, Switzerland, KPMG International.
[43] Chaka, M. (2018). Sustainability reporting in South Africa: A descriptive and comparative analysis, *Journal of Gender, Information and Development in Africa,* Vol. 7, No. 2, 159-184, doi: 10.31920/2050-4284/2018/v7n2a8.

[44] Grewal, J., Serafeim, G., Yoon, A. (2016). Shareholder activism on sustainability issues, *Harvard Business School Working paper*, No. 17-003, doi: 10.2139/ssrn.2805512.
[45] Nugroho, D.M. (2015). Pengaruh kualitas pelayanan, kualitas produk layanan, dan harga produk layanan terhadap kepuasan pelanggan serta dampaknya terhadap loyalitas pelanggan prabayar telkomsel, *Jurnal Operations Excellence: Journal of Applied Industrial Engineering,* Vol. 7, No. 2, 158-174.
[46] Colwell, S., Hogarth-Scott, S., Jiang, D., Joshi, A. (2009). Effects of organizational and serviceperson orientation on customer loyalty, *Management Decision*, Vol. 47, No. 10, 1489-1513, doi: 10.1108/00251740911004655.
[47] Marakanon, L., Panjakajornsak, V. (2017). Perceived quality, perceived risk and customer trust affecting customer loyalty of environmentally friendly electronics products, *Kasetsart Journal of Social Sciences,* Vol. 38, No. 1, 24-30, doi: 10.1016/j.kjss.2016.08.012.
[48] Kitchenham, B., Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE 2007-001), Technical report, Keele University and Durham University Joint Report.
[49] Bolderston, A. (2008). Writing an effective literature review, *Journal of Medical Imaging and Radiation Sciences*, Vol. 39, No. 2, 86-92, doi: 10.1016/j.jmir.2008.04.009.
[50] Kumar, C.S.C., Naidu, N.V.R. (2012). Minimizing the eighth waste of lean-absenteeism through six sigma methodology, *International Journal for Quality Research*, Vol. 6, No. 2, 113-118.
[51] Naidu, N.V.R. (2011). Development of mathematical models for optimal preventive maintenance policy in a steel industry: Six sigma approach, *International Journal for Quality research,* Vol. 5, No. 3, 179-185.
[52] Zhan, W. (2008). A six sigma approach for the robust design of motor speed control using modelling and simulation, *International Journal of Six Sigma and Competitive Advantage,* Vol. 4, No. 2, 95-113, doi: 10.1504/ IJSSCA.2008.020277.
[53] Gajbhiye, P.R., Waghmare, A.C., Parikh, R.H. (2016). Applications of lean six sigma methodologies for improvement in industrial safety, *International Journal for Innovative Research in Science & Technology*, Vol. 3, No. 6, 168-175.
[54] Chang, S.-I., Yen, D.C., Chou, C.-C., Wu, H.-C., Lee, H.-P. (2012). Applying six sigma to the management and improvement of production planning procedure's performance, *Total Quality Management & Business Excellence,* Vol. 23, No. 3-4, 291-308, doi: 10.1080/14783363.2012.657387.
[55] Purnama, G.N.T., Gunanto, Sugengriadir, R.M. (2019). Increasing the implementation of environmental management systems based ISO 14001 with the six sigma: Case study method in a manufacturing industry, *International Journal of Research in Engineering, Science and Management*, Vol. 2, No. 1, 103-106.
[56] El Hassani, I., Benlaajili, S., Nokra, N. (2017). Optimization of the particle size distribution of white sugar using six sigma methodology, *International Journal of Advance Engineering and Research Development*, Vol. 4, No. 9, 190-204, doi: 10.21090/IJAERD.71557.
[57] Soković, M., Pavletić, D., Krulčić, E. (2006). Six sigma process improvements in automotive parts production, *Journal of Achievements in Materials and Manufacturing Engineering,* Vol. 19, No. 1, 96-102.
[58] Rathilall, R., Singh, S. (2018). A lean six sigma framework to enhance the competitiveness in selected automotive component manufacturing organisations, *South African Journal of Economic and Management Sciences*, Vol. 21, No. 1, 1-13, doi: 10.4102/sajems.v21i1.1852.
[59] Kaushik, P., Khanduja, D. (2007). DM make up water reduction in thermal power plants using six sigma DMAIC methodology, *Journal of Scientific & Industrial Research*, Vol. 67, 36-42.
[60] Desai, D., Prajapati, B.N. (2017). Competitive advantage through Six Sigma at plastic injection molded parts manufacturing unit: A case study, *International Journal of Lean Six Sigma*, Vol. 8, No. 4, 411-435, doi: 10.1108/IJLSS-06-2016-0022.
[61] Gijo, E.V., Scaria, J., Antony, J. (2011). Application of six sigma methodology to reduce defects of a grinding process, *Quality and Reliability Engineering International,* Vol. 27, No. 8, 1221-1234, doi: 10.1002/qre.1212.
[62] Anderson, N.C., Kovach, J.V. (2014). Reducing welding defects in turnaround projects: A lean six sigma case study, *Quality Engineering*, Vol. 26, No. 2, 168-181, doi: 10.1080/08982112.2013.801492.
[63] Sharma, P., Malik, S.C., Gupta, A., Jha, P.C. (2018). A DMAIC six sigma approach to quality improvement in the anodizing stage of the amplifier production process, *International Journal of Quality & Reliability Management,* Vol. 35, No. 9, 1868-1880, doi: 10.1108/IJQRM-08-2017-0155.
[64] Kumaradivel, A., Natarajan, U. (2013). Application of six sigma DMAIC methodology to sand-casting process with response surface methodology, *The International Journal of Advanced Manufacturing Technology,* Vol. 69, 1403-1420, doi: 10.1007/s00170-013-5119-2.
[65] Lo, W.C., Tsai, K.M., Hsieh, C.Y. (2009). Six sigma approach to improve surface precision of optical lenses in the injection-molding process*, The International Journal of Advanced Manufacturing Technology*, Vol. 41, 885-896, doi: 10.1007/s00170-008-1543-0.
[66] Chen, J.C., Li, Y., Cox, R.A. (2009). Taguchi-based six sigma approach to optimize plasma cutting process: An industrial case study, *The International Journal of Advanced Manufacturing Technology,* Vol. 41, 760-769, doi: 10.1007/s00170-008-1526-1.
[67] Jacob, J.J., Jenson J.E. (2014). Reducing the throughput time by value stream mapping in a tyre manufacturing industry, *International Journal of Engineering Research & Technology*, Vol. 3, No. 11, 385-388.

APEM journal

Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 326–334
https://doi.org/10.14743/apem2021.3.403

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

Modeling and optimization of finish diamond turning of spherical surfaces based on response surface methodology and cuckoo search algorithm

Kramar, D.[a,*], **Cica, Dj.**[b]

[a]University in Ljubljana, Faculty of Mechanical Engineering Ljubljana, Slovenia
[b]University in Banja Luka, Faculty of Mechanical Engineering Banja Luka, Bosnia and Herzegovina

A B S T R A C T

Surface roughness is one of the most significant factors to indicate the product quality. Diamond turning is an efficient and highly accurate material removal process to improve the surface quality of the workpiece. In the present study, the arithmetic mean absolute roughness (R_a) and total height of profile (R_t) of spherical surface during finish turning of a commercial brass alloy CuZn40Pb2 were modeled using Response Surface Methodology (RSM). The experimental investigations were carried out using the Central Composite Design (CCD) under dry conditions. The effect of cutting parameters such as spindle speed, feed rate and depth of cut) on spherical surface quality was analyzed using analysis of variance (ANOVA). A cuckoo search (CS) algorithm was used to determine the optimum machining parameters to minimize the surface roughness. Finally, confirmation experiments were carried out to verify the adequacy of the considered optimization algorithm.

A R T I C L E   I N F O

Keywords:
Brass alloy;
Diamond turning;
Surface roughness;
Spherical surface;
Modeling;
Optimization;
Response surface methodology (RSM);
Cuckoo search (CS)

**Corresponding author:*
davorin.kramar@fs.uni-lj.si
(Kramar, D.)

Article history:
Received 31 May 2021
Revised 3 September 2021
Accepted 15 September 2021

1. Introduction

Brass (copper-zinc) alloys are used in a wide range of industrial applications such as mechanical, electrical and hydraulic systems due to their excellent formability, high thermal and electrical conductivity, corrosion resistance and excellent antibacterial properties. To improve machinability, various alloying elements are commonly added to brass. The most important of these elements is lead, which improves machinability in terms of tool wear, cutting forces, chip breaking and surface quality [1-3]. Extensive research has been carried out in recent decades to improve the machinability of brass alloys.

In [4], a comparison of the machinability of three lead-free brass alloys and one leaded brass alloy in terms of energy consumption and chip morphology was carried out. The influence of different coating types and the use of diamond tools on cutting forces, chip formation and surface quality was investigated by Klocke *et al.* [5]. Nobel *et al.* [6] analyzed the chip formation process in different low lead brass alloys. Schultheiss *et al.* [7] analyzed cutting forces, surface quality and tool wear during longitudinal turning to evaluate the machinability of leaded and lead-free brass alloys. The effects of minimum quantity lubrication (MQL) and cutting parameters on surface quality during turning of commercial brass were studied by Davim *et al.* [8]. A comparison with conventional flooding conditions was also made. The machinability of the highly leaded brass alloy CuZn39Pb3 [9] and commercial lead-free brass alloys [10] was also analyzed. Vilarinho *et al.* [11] studied the influence of the chemical composition of brass alloys on surface quality and machining forces during turning. Schultheiss *et al.* [12] compared the machinability and manufacturing costs in turning of conventional leaded brass alloys and a low-lead alternative. Toulfatzis *et al.* [13] studied the chip morphology and tool wear during longitudinal turning of two leaded brass alloys. The Taguchi method with the utility concept was introduced for simultaneous optimization of surface quality and specific cutting force in turning of brass CuZn39Pb3 under MQL cutting [14]. Toulfatzis *et al.* [15] used the signal-to-noise ratio for single optimization of surface roughness and cutting force in longitudinal turning of lead-free and leaded brass alloys.

Some researchers also used artificial intelligence based methods. Gaitonde *et al.* [16] applied a genetic algorithm to determine the optimum machining parameters for minimizing surface roughness when turning leaded brass alloys under MQL conditions. Raja and Baskar [17] presented a multi-objective optimization method based on particle swarm optimization with two objectives, namely machining time and surface quality. Optimum cutting conditions during turning for different materials namely brass, aluminum, mild steel and copper were selected. Natarajan *et al.* [18] estimated the surface roughness in longitudinal turning of brass using artificial neural networks.

After a review of the literature, it is apparent that a considerable amount of research has been conducted on turning of brass alloys. However, none of the literature reviewed dealt with finish diamond turning of spherical surfaces. As market competitiveness increases, surface roughness is probably one of the most widely used indicators of surface quality of machined parts today [19-27]. Surface roughness affects several functional properties of products, especially friction, fatigue strength, wear, heat transfer, light reflection, lubricant distribution, etc. Ensuring surface quality is one of the most critical issues in the fully automated mass production of parts with spherical surfaces. However, surface roughness is strongly influenced by the variation of process parameters and analytical modeling is difficult due to its nonlinearity. Therefore, it is imperative to develop a mathematical model of surface roughness, exploit the influence of machining parameters, and finally optimize the surface quality in brass ball turning.

2. Experimental procedure

The machining tests were performed on the special machine tool Picchi Diamantatrice to produce spherical parts with a spindle power of 5.5 kW and a maximum spindle speed of 7000 rpm. This lathe was equipped with two tools: a carbide turning tool for roughing and a diamond tool for finishing and machines spheres with a diameter of 2" (Fig. 1). Commercially available brass CuZn40Pb2 (CW617N) workpieces were used for machining. The chemical composition of the material is summarized in Table 1. The working material has a hardness of 90-100 HRB and a tensile strength of 390-440 N/mm^2. This alloy provides good corrosion resistance together with good formability.

Table 1 Chemical composition of the studied brass alloys

Alloy	Cu (%)	Zn (%)	Pb (%)	Al (%)	Fe (%)	Ni (%)	Sn (%)	Others: Mn, Si, Sb, As, Bi, (%)
CuZn40Pb2	57-59	rest	1.6-2.5	0.05 max	0.3 max	0.3 max	0.3 max	0.2 max

The diamond cutting tool was used for precision machining of a ball valve. The tool nose does not have a radius like conventional cutting tools, but a flat nose that allows very fine machining. The dimensions of the tool were 95 × 12 × 12 mm, with the sides at a 45° angle and a 1.3 mm wide cutting edge. The main angles of the tool cutting wedge were 2° for clearance angle α, -10.5° for rake angle γ and consequently 98.5° for wedge angle β. The cutting depth of the diamond segment is 0.4 mm and represents the maximum theoretical cutting depth, but due to the possibility of pulling the diamond segment out of the tool holder, the depth of cut is usually not used deeper than 0.2 mm.

Fig. 1 Experimental setup for diamond fine turning of CuZn40Pb2

The arithmetic mean of the absolute roughness (R_a) and a total height of profile (R_t) of the machined workpiece were measured with the surface roughness measuring device Tesa Rugosurf 20 (Fig. 2). The examined length was 3.2 mm with a basic span of 0.8 mm.

Fig. 2 Surface roughness measurements

In the present study, three cutting parameters, i.e., spindle speed (n), feed rate (f) and depth of cut (a) were selected, and the ranges of machining conditions were defined through initial experiments. The cutting parameters and their levels are given in Table 2.

The experiments were designed and carried out according to the Central Composite Design (CCD). All experiments were conducted under dry conditions. The measured values of R_a and R_t are shown in Table 3.

Table 2 Machining parameters and their levels

Parameter	Unit	Levels		
		Level 1	Level 2	Level 3
Spindle speed, n	min^{-1}	4800	5500	6200
Feed, f	mm/rev	0.48	0.64	0.80
Depth of cut, a	mm	0.075	0.125	0.175

Table 3 Experimental layout for central composite design

No.	Process parameters			Response measurements	
	n (min^{-1})	f (mm/rev)	a (mm)	R_a (μm)	R_t (μm)
1	5500	0.64	0.125	0.140	0.97
2	4800	0.48	0.075	0.208	1.38
3	6200	0.80	0.075	0.164	1.11
4	5500	0.48	0.125	0.173	1.18
5	4800	0.48	0.175	0.212	1.76
6	4800	0.80	0.075	0.249	2.10
7	5500	0.64	0.125	0.154	1.00
8	6200	0.64	0.125	0.155	1.07
9	5500	0.80	0.125	0.152	1.11
10	4800	0.64	0.125	0.179	1.29
11	5500	0.64	0.125	0.156	1.06
12	5500	0.64	0.075	0.149	1.05
13	4800	0.80	0.175	0.224	1.56
14	6200	0.80	0.175	0.150	1.14
15	6200	0.48	0.175	0.215	1.56
16	6200	0.48	0.075	0.145	1.35
17	5500	0.64	0.175	0.218	1.51

3. Results and discussion

The reduced quadratic polynomial model, given as a function of the machining parameters for an arithmetic mean of the absolute roughness (R_a) and a total height of the profile (R_t) using the Response Surface Methodology (RSM) is represented by equations (1) and (2), respectively

$$R_a = 0.4063 - 3.4714 \cdot 10^{-5} n + 0.212 f - 2.1591 a - 1.7656 f a + 13.9886 a^2 \quad (1)$$

$$R_t = 8.5175 - 3.2011 \cdot 10^{-3} n + 9.2605 f - 14.7502 a - 1.317 \cdot 10^{-3} n f - 17.1875 f a + 3.4347 \cdot 10^{-7} n^2 + 107.3208 a^2 \quad (2)$$

The adequacy of the developed models was examined using analysis of variance (ANOVA). The results are given in Table 4 and Table 5 for R_a and R_t, respectively. The F-values for R_a and R_t are 5.80 and 7.85, respectively. For the desired confidence level (95 %), the F-values of the established model exceed the F-value of the standard tabulated F-values for R_a and R_t of 3.204 and 3.293, respectively. Thus, the two reduced quadratic models can be considered appropriate within the confidence limit. P-values smaller than 0.05 imply that n and the product a^2 are statistically significant terms for the arithmetic mean of absolute roughness R_a, while n and the products $n \times f$, $f \times a$ and a^2 are also observed to be significant terms for the total height of the profile R_t. Other parameters have no statistical significance for the surface quality responses considered. The squared correlation coefficient (R^2) values of 0.7250 and 0.8593 for R_a and R_t, respectively, show good agreement among the experimental and predicted values for both models.

The 3D response surfaces were also created to study the effects of the machining parameters and their interactions. Figs. 3-5 show the contour plots for the surface roughness parameters (R_a and R_t) as a function of spindle speed (n), feed rate (f) and depth of cut (a).

Table 4 The ANOVA table for R_a

Source	Sum of squares	DF	Mean square	F-value	P-value
Model	0.014	5	$2.728 \cdot 10^{-3}$	5.80	0.0073
n	$5.905 \cdot 10^{-3}$	1	$5.905 \cdot 10^{-3}$	12.56	0.0046
f	$1.96 \cdot 10^{-5}$	1	$1.96 \cdot 10^{-5}$	0.042	0.8419
a	$1.082 \cdot 10^{-3}$	1	$1.082 \cdot 10^{-3}$	2.30	0.1575
$f \times a$	$1.596 \cdot 10^{-3}$	1	$1.596 \cdot 10^{-3}$	3.39	0.0925
a^2	$5.036 \cdot 10^{-3}$	1	$5.036 \cdot 10^{-3}$	10.71	0.0074
Error	$5.172 \cdot 10^{-3}$	11	$4.072 \cdot 10^{-4}$		
Total	0.019	16			

Table 5 The ANOVA table for R_t

Source	Sum of squares	DF	Mean square	*F*-value	*P*-value
Model	1.31	7	0.19	7.85	0.0031
n	0.35	1	0.35	14.53	0.0041
f	$4.41\cdot10^{-3}$	1	$4.41\cdot10^{-3}$	0.19	0.677
a	0.029	1	0.029	1.22	0.2971
n×f	0.17	1	0.17	7.31	0.0242
f×a	0.15	1	0.15	6.35	0.0327
n^2	0.086	1	0.086	3.60	0.0902
a^2	0.22	1	0.22	9.16	0.0143
Error	0.21	9	0.024		
Total	1.52	16			

The effects of spindle speed and feed rate on roughness parameters are shown in Fig. 3, where the depth of cut is kept at an intermediate level. The minimum value of R_a was obtained at high spindle speed and low feed rate, as shown in Fig. 3a. It should also be noted that the R_a value is almost directly proportional to these two parameters, with feed rate having less significance. Fig. 3b shows that R_t is significantly affected by spindle speed, while feed rate has less influence on it. The minimum value of R_t was also found at the maximum values of spindle speed and feed rate. Therefore, to obtain better roughness parameters, higher spindle speeds and feed rate should be preferred for diamond finish turning of spherical surfaces.

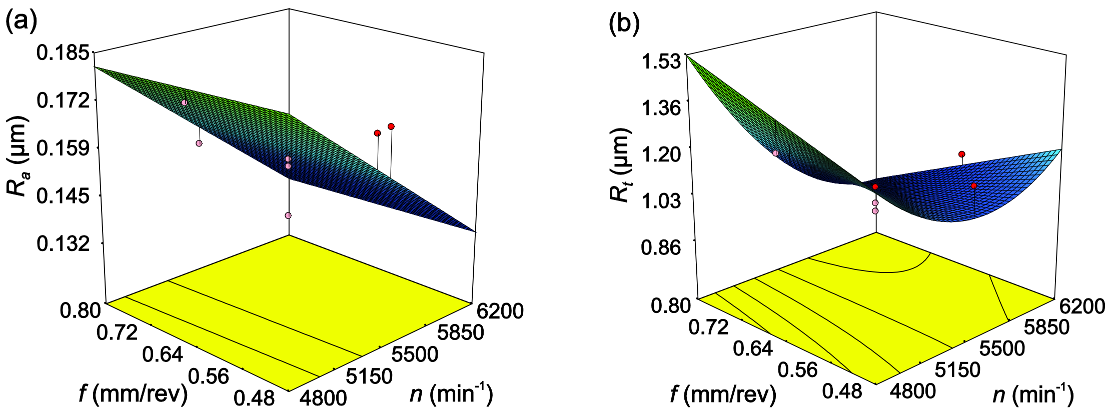


Fig. 3 Surface plot of R_a (a) and R_t (b) with spindle speed and feed rate

Fig. 4 shows the estimated response surface with respect to the spindle speed and the depth of cut, while the feed rate is kept at an intermediate level. The minimum values of both parameters were obtained at high spindle speed and medium depth of cut. It can also be observed that both cutting parameters strongly influence the surface quality.

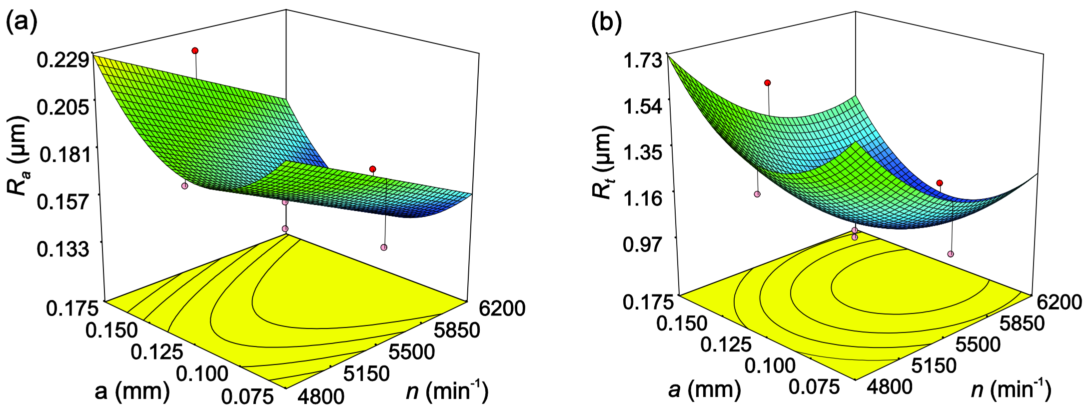


Fig. 4 Surface plot of R_a (a) and R_t (b) with spindle speed and depth of cut

Fig. 5 shows the influences of feed rate and depth of cut on the surface roughness parameters (R_a and R_t), while the spindle speed is kept at an intermediate level. The plots show that the minimum values for both parameters were found at a high feed rate, while the depth of cut was about 0.1 mm. This figure shows that both roughness parameters are almost directly proportional to the feed rate. It is also worth noting that the depth of cut is more significant compared to the feed rate.

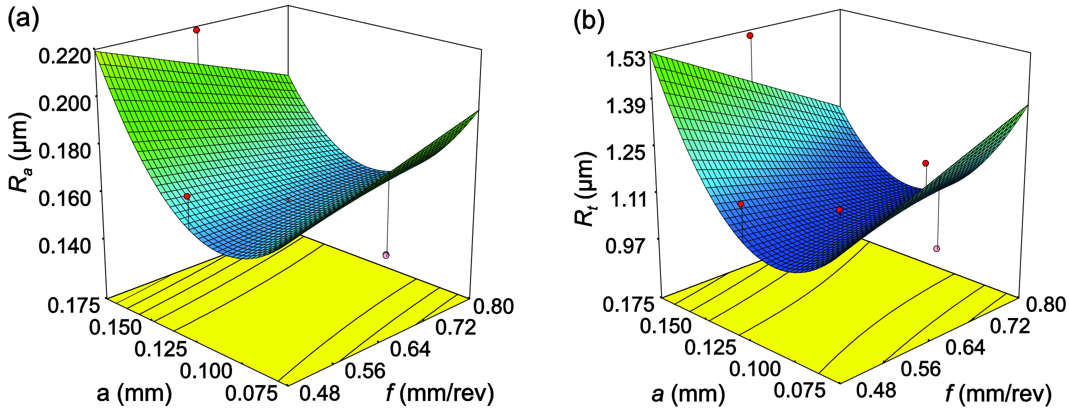


Fig. 5 Surface plot of R_a (a) and R_t (b) with feed rate and depth of cut

4. Optimization

4.1 Cuckoo search

The cuckoo search (CS) algorithm, originally introduced by Yang and Deb [28], is one of the recent swarm intelligence-based optimization algorithms. The CS algorithm was inspired by the brood parasitism behavior of certain cuckoo bird species and the characteristics of Lévy flights discovered in the flight behavior of numerous insects and animals. Female cuckoos lay their eggs in the nests of other, typically different species of host birds. These eggs also resemble the host birds' eggs in color and pattern. When the host bird realizes that the eggs are not its own, it either discards them or simply leaves the nest and builds a new one elsewhere. Cuckoos must therefore mimic their host birds' eggs very closely to minimize the likelihood that their eggs will be abandoned. Optimization is about replacing a less good nesting solution with a new and potentially better one (cuckoo).

The CS algorithm is essentially based on three idealized rules [29]: (i) each cuckoo lays one egg at a time in a randomly selected nest; (ii) an elite selection procedure is used in which only the best nests with superior quality eggs are passed on to the next generation; (iii) the number of available host nests is fixed and a host bird can detect a foreign egg with probability in the range of [0, 1].

While producing new solutions $x^{(t+1)}$ for the i-th cuckoo, Lévy flight is performed

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \otimes \text{Lévy}(\lambda) \quad (3)$$

where $x^{(t)}_i$ denotes the i-th candidate solution at iteration t and $\alpha > 0$ denotes the step size factor. In most cases, $\alpha = 1$ can be used. The product $\otimes$ denotes inputwise walking for multiplications.

Basically, the Lévy flight results in a random walk, where the random step length is determined by a Lévy distribution with infinite mean and infinite variance

$$\text{Lévy} \sim u = t^{-\lambda}, (1 < \lambda \leq 3) \quad (4)$$

4.2 Optimization model

The optimization process in this paper aims to find the best combination of process parameter levels that results in the lowest value of R_a and R_t. To formulate the optimization problem, the regression models for R_a (Eq. 1) and R_t (Eq. 2) were used as the objective function. In this study, three variables, namely spindle speed, feed rate and depth of cut were considered as optimization variables. Normalization of each subobjective was also introduced to compensate the differences in numerical values between them. Thus, the resulting objective function (ROF) to be minimized is a weighted combination of the two objectives as follows:

$$\mathrm{ROF}\ (n, f, a) = w_1 \frac{R_a}{R_{a\min}} + w_2 \frac{R_t}{R_{t\min}} \tag{5}$$

where w_1 and w_2 are the weight factors of R_a and R_t, respectively. In this paper, equal weights for R_a and R_t were selected, i.e. $w_1 = w_2 = 1/2$.

The minimization of the ROF (Eq. 5) is subject to the limits of the cutting parameters. The boundary conditions were the upper and lower limit of the experimental machining parameters (Table 2) and were given as follows: $4800 \leq n \leq 6200$ min^{-1}, $0.48 \leq f \leq 0.8$ mm/rev, $0.075 \leq a \leq 0.175$ mm.

After formulating the optimization problem and its constraints, the CS optimization algorithm was employed to solve the problem. The proposed CS algorithm requires some setting parameters for implementation. The minimum value of the resulting objective function (ROF = 0.916) was obtained for a population size of 20, termination probability of 0.25, and termination criterion of 100 generations. The results of the CS optimization algorithm showed that the best combination of turning parameters for simultaneous optimization of the arithmetic mean of absolute roughness (R_a) and total height of profile (R_t) was: 6200 min^{-1} for spindle speed, 0.8 mm/rev for feed rate, and 0.13 for depth of cut.

4.3 Confirmation experiments

The confirmation tests were carried out at the optimum values of the process parameters to verify the quality characteristics of the spherical fine turning process recommended in the study. In accordance with the obtained optimum results, four new experiments were carried out (Table 6). The mean values of R_a and R_t were 0.139 µm and 1.04 µm, respectively, which were in good agreement with the predicted values. Consequently, the proposed CS optimization algorithm was efficient to find out the optimal set of machining parameters for spherical finish turning associated with minimum surface roughness.

Table 6 Results of confirmation tests

Confirmation test	1	2	3	4	Mean
R_a (µm)	0.143	0.149	0.134	0.141	0.139
R_t (µm)	1.19	0.89	1.10	0.97	1.04

5. Conclusions

In the present research work, the influence of machining parameters such as spindle speed, feed rate and depth of cut on the arithmetic mean of absolute roughness (R_a) and total height of profile (R_t) in diamond finish turning of brass CuZn40Pb2 was investigated. The following conclusions can be drawn from the analysis of the results, subsequent model development and optimization:

- The response surface methodology in combination with the central composite design has been successfully applied to study the effects of different machining parameters on two quality characteristics, namely the arithmetic mean of the absolute surface roughness and the total profile height, during diamond finish turning of spherical surfaces. The developed mathematical models of both responses in terms of the actual design factors, their interactions and quadratic terms are suitable for the analysis of the sphere turning process.
- The spindle speed and the quadratic term of the depth of cut were the significant parameters affecting the arithmetic mean of the absolute roughness, according to the results of ANOVA. The ANOVA also showed that the spindle speed, the quadratic term of the depth of cut, the interaction terms between the spindle speed and the feed rate, and the feed

rate and the depth of cut were significant terms affecting the performance of the total height of the profile. Spindle speed showed the greatest influence on both surface parameters compared to depth of cut and feed rate.

- The best combination of cutting parameters to simultaneously minimize the arithmetic mean of the absolute roughness and the total height of the profile, obtained by the optimization model based on the cuckoo search algorithm, was as follows: 6200 min-1 for spindle speed, 0.8 mm/rev for feed rate and 0.13 mm for depth of cut. This study and the determined combination of parameters allow a more detailed technology planning for the production of high-quality spherical components, both from the point of view of premachining and productivity itself. The results obtained do not confirm the theory of turning. In fact, in our case, we obtain the best results at the highest feed rate, which is contrary to expectations. This is certainly related to the kinematics and stability of the machine tool in this speed and feed range, which in turn shows how important it is to optimize each machining process.
- The results of the confirmation experiments reveal the effectiveness of the proposed algorithm for optimizing the surface quality characteristics in diamond finish turning of spherical surfaces.

Acknowledgement

The research was conducted within the framework of a scientific and technological intergovernmental bilateral project supported by the Slovenian Research Agency under the project serial numbers: P2-0266 and BI-BA/19-20-008, and by the relevant ministries Government of the Republic of Srpska, Bosnia and Herzegovina.

References

[1] Kuyucak, S., Sahoo, M. (1996). A review of the machinability of copper-base alloys, *Canadian Metallurgical Quarterly*, Vol. 35, No. 1, 1-15, doi: 10.1016/0008-4433(95)00023-2.

[2] Chen, P.-C., Chen, Y.-C., Pan, C.-W., Li, K.-M. (2015). Parameter optimization of micromilling brass mold inserts for microchannels with Taguchi method, *International Journal of Precision Engineering and Manufacturing*, Vol. 16, 647-651, doi: 10.1007/s12541-015-0086-1.

[3] Pantazopoulos, G. (2002). Leaded brass rods C 38500 for automatic machining operations: A technical report, *Journal of Materials Engineering and Performance*, Vol. 11, No. 4, 402-407, doi: 10.1361/105994902770343926.

[4] Toulfatzis, A.I., Pantazopoulos, G.A., Besseris, G.J., Paipetis, A.S. (2016). Machinability evaluation and screening of leaded and lead-free brasses using a non-linear robust multifactorial profiler, *The International Journal of Advanced Manufacturing Technology*, Vol. 86, No. 9-12, 3241-3254, doi: 10.1007/s00170-016-8435-5.

[5] Klocke, F., Nobel, C., Veselovac, D. (2016). Influence of tool coating, tool material, and cutting speed on the machinability of low-leaded brass alloys in turning, *Materials and Manufacturing Processes*, Vol. 31, No. 14, 1895-1903, doi: 10.1080/10426914.2015.1127944.

[6] Nobel, C., Hofmann, U., Klocke, F., Veselovac, D. (2016). Experimental investigation of chip formation, flow, and breakage in free orthogonal cutting of copper-zinc alloys, *The International Journal of Advanced Manufacturing Technology*, Vol. 84, No. 5-8, 1127-1140, doi: 10.1007/s00170-015-7749-z.

[7] Schultheiss, F., Johansson, D., Bushlya, V., Zhou, J., Nilsson, K., Ståhl, J-E. (2017). Comparative study on the machinability of lead-free brass, *Journal of Cleaner Production*, Vol. 149, 366-377, doi: 10.1016/j.jclepro.2017.02.098.

[8] Davim, J.P., Sreejith, P.S., Silva, J. (2007). Turning of brasses using minimum quantity of lubricant (MQL) and flooded lubricant conditions, *Materials and Manufacturing Processes*, Vol. 22, No. 1, 45-50, doi: 10.1080/10426910601015881.

[9] Vaxevanidis, N.M., Fountas, N.A., Koutsomichalis, A., Kechagias, J.D. (2018). Experimental investigation of machinability parameters in turning of CuZn39Pb3 brass alloy, *Procedia Structural Integrity*, Vol. 10, 333-341, doi: 10.1016/j.prostr.2018.09.046.

[10] Nobel, C., Klocke, F., Lung, D., Wolf, S. (2014). Machinability enhancement of lead-free brass alloys, *Procedia CIRP*, Vol. 14, 95-100, doi: 10.1016/j.procir.2014.03.018.

[11] Vilarinho, C., Davim, J.P., Soares, D., Castro, F., Barbosa, J. (2005). Influence of the chemical composition on the machinability of brasses, *Journal of Materials Processing Technology*, Vol. 170, No. 1-2, 441-447, doi: 10.1016/j.jmatprotec.2005.05.035.

[12] Schultheiss, F., Windmark, C., Sjöstrand, S., Rasmusson, M., Ståhl, J-E. (2018). Machinability and manufacturing cost in low-lead brass, *The International Journal of Advanced Manufacturing Technology*, Vol. 99, No. 9-12, 2101-2110, doi: 10.1007/s00170-018-1866-4.

[13] Toulfatzis, A.I., Besseris, G.J., Pantazopoulos, G.A., Stergiou, C. (2011). Characterization and comparative machinability investigation of extruded and drawn copper alloys using non-parametric multi-response optimization and orthogonal arrays, *The International Journal of Advanced Manufacturing Technology*, Vol. 57, No. 5-8, 811-826, doi: 10.1007/s00170-011-3319-1.

[14] Gaitonde, V.N., Karnik, S.R., Davim, J.P. (2008). Selection of optimal MQL and cutting conditions for enhancing machinability in turning of brass, *Journal of Materials Processing Technology*, Vol. 204, No. 1-3, 459-464, doi: 10.1016/j.jmatprotec.2007.11.193.

[15] Toulfatzis, A.I., Pantazopoulos, G.A., David, C.N., Sagris, D.S., Paipetis, A.S. (2018). Machinability of eco-friendly lead-free brass alloys: Cutting-force and surface-roughness optimization, *Metals*, Vol. 8, No. 4, Article No. 250, doi: 10.3390/met8040250.

[16] Gaitonde, V.N., Karnik, S.R., Davim, J.P. (2012). Optimal MQL and cutting conditions determination for desired surface roughness in turning of brass using genetic algorithms, *Machining Science and Technology*, Vol. 16, No. 2, 304-320, doi: 10.1080/10910344.2012.673976.

[17] Raja, S.B., Baskar, N. (2011). Particle swarm optimization technique for determining optimal machining parameters of different work piece materials in turning operation, *The International Journal of Advanced Manufacturing Technology*, Vol. 54, No. 5-8, 445-463, doi: 10.1007/s00170-010-2958-y.

[18] Natarajan, C., Muthu, S., Karuppuswamy, P. (2011). Prediction and analysis of surface roughness characteristics of a non-ferrous material using ANN in CNC turning, *The International Journal of Advanced Manufacturing Technology*, Vol. 57, No. 9-12, 1043-1051, doi: 10.1007/s00170-011-3343-1.

[19] Brezocnik, M., Kovacic, M. (2003). Integrated genetic programming and genetic algorithm approach to predict surface roughness, *Materials and Manufacturing Processes*, Vol. 18, No. 3, 475-491, doi: 10.1081/AMP-12002 2023.

[20] Sekulic, M., Pejic, V., Brezocnik, M., Gostimirović, M., Hadzistevic, M. (2018). Prediction of surface roughness in the ball-end milling process using response surface methodology, genetic algorithms, and grey wolf optimizer algorithm, *Advances in Production Engineering & Management*, Vol. 13, No. 1, 18-30, doi: 10.14743/apem2018.1. 270.

[22] Klancnik, S., Hrelja, M., Balic, J., Brezocnik, M. (2016). Multi-objective optimization of the turning process using a Gravitational Search Algorithm (GSA) and NSGA-II approach, *Advances in Production Engineering & Management*, Vol. 11, No. 4, 366-376, doi: 10.14743/apem2016.4.234.

[22] Brezocnik, M., Kovacic, M., Ficko, M. (2004). Prediction of surface roughness with genetic programming, *Journal of Materials Processing Technology*, Vol. 157-158, 28-36, doi: 10.1016/j.jmatprotec.2004.09.004.

[23] Sung, A.N., Loh, W.P., Ratnam, M.M. (2016). Simulation approach for surface roughness interval prediction in finish turning, *International Journal of Simulation Modelling*, Vol. 15, No. 1, 42-55, doi: 10.2507/IJSIMM15(1) 4.320.

[24] Özdemir, M. (2020). Modelling and prediction of effect of machining parameters on surface roughness in turning operations, *Tehnički Vjesnik – Technical Gazette*, Vol. 27, No. 3, 751-760, doi: 10.17559/TV-20190320104114.

[25] Boyacı, A.İ., Hatipoğlu, T., Balcı, E. (2017). Drilling process optimization by using fuzzy-based multi-response surface methodology, *Advances in Production Engineering & Management*, Vol. 12, No. 2, 163-172, doi: 10.14743/apem2017.2.248.

[26] Sterpin Valic, G., Cukor, G., Jurkovic, Z., Brezocnik, M. (2019). Multi-criteria optimization of turning of martensitic stainless steel for sustainability, *International Journal of Simulation Modelling*, Vol. 18, No. 4, 632-642, doi: 10.2507/IJSIMM18(4)495.

[27] Aljinović, A., Bilić, B., Gjeldum, N., Mladineo, M. (2021). Prediction of surface roughness and power in turning process using response surface method and ANN, *Tehnički Vjesnik – Technical Gazette*, Vol. 28, No. 2, 456-464, doi: 10.17559/TV-20190522104029.

[28] Yang, X.-S., Deb, S. (2009). Cuckoo search via Lévy flights, In: *Proceedings of 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC),* Coimbatore, India, 210-214, doi: 10.1109/NABIC.2009.5393690.

[29] Yang, X.-S., Deb, S. (2014). Cuckoo search: Recent advances and applications, *Neural Computing and Applications*, Vol. 24, No. 1, 169-174, doi: 10.1007/s00521-013-1367-1.

APEM journal

Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 335–347
https://doi.org/10.14743/apem2021.3.404

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

Tactical manufacturing capacity planning based on discrete event simulation and throughput accounting: A case study of medium sized production enterprise

Jurczyk-Bunkowska, M.[a,*]

[a]Faculty of Economics and Management, Opole University of Technology, Poland

A B S T R A C T

The article presents the application of the original methodology to support tactical capacity planning in a medium-sized manufacturing company. Its essence is to support medium-term decisions regarding the development of the production system through economic assessment of potential change scenarios. It has been assumed that the developed methodology should be adapted to small and medium-sized enterprises (SMEs). Due to their flexibility, they usually have limited time for decision-making, and due to limited financial resources, they rely on internal competencies. The proposed approach that does not require mastery of mathematical modelling but allows streamlining capacity planning decisions. It uses the reasoning of throughput accounting (TA) supported by data obtained based on discrete event simulation (DES). Using these related tools in the design and analysis of change scenarios, make it possible for SME managers to make a rational decision regarding the development of the production system. Case studies conducted in a roof window manufacturing company showed the methodology. The application example presented in the article includes seven change scenarios analyzed based on computer simulations by the software Tecnomatix Plant Simulation. The implementation of the approach under real conditions has shown that a rational decision-making process is possible over time scale and with the resources available to SMEs for this type of decision.

A R T I C L E I N F O

Keywords:
Decision process;
Capacity planning;
Discrete event simulation (DES);
Throughput accounting (TA);
Plant simulation;
Small and medium-sized enterprises (SME);
Production scenarios;
Tecnomatix Plant Simulation

**Corresponding author:*
m.jurczyk-bunkowska@po.edu.pl
(Jurczyk-Bunkowska, M.)

Article history:
Received 7 June 2021
Revised 12 October 2021
Accepted 13 October 2021

1. Introduction

Balancing supply-to-demand production capacity is a key consideration to ensure the viability of the company's operating activities. The need for changes in the manufacturing system is dictated by adapting to increasingly precisely projected demand [1], the increasing possibility of automating and robotizing monotonous work requiring physical exertion [2], implementing technologies in the field of Industry 4.0 that significantly improve production processes [3]. Rational implementation of changes requires an understanding of their consequences. The right decisions allow maintaining the right balance between system capacity and production costs. Production capacity is the most strategic internal capability that manufacturing firms must create, sustain and plan for. Capacity management aims to ensure that a manufacturer has the 'right' capacity to act within a complex structure and how best to 'utilise' their internal capabilities [4]. They refer to the amount of output that a system is capable of achieving over a specific period of time [5].

Both capacity expansion and contraction are called capacity planning [6]. The level of production capacity directly influences on the one hand the possibility of carrying out orders and on the other hand the level of fixed costs of the company. Therefore, the cost of production is determined by the level of capacity utilisation and thus its match with demand [7]. With regard to the medium - term capacity planning, appropriate capacity planning methods are key, which assess the consequences of these decisions on the production system [8].

Observations made in small and medium-sized manufacturing companies indicate managers are under pressure to ensure that the decision-making process for tactical capacity planning does not exceed a few weeks. It includes an analysis of the production situation leading to the development of change scenarios and their assessment, which results in a recommendation for investments. These are issues so complex that leaving their intuition even an experienced manager is irrational. The literature on the problem of capacity planning offers solutions based on mathematical modelling and/or artificial intelligence. There are lack of approaches that would ensure a rational solution to the problem, even with a reduction in its optimization. As part of the presented research, it sought a method that would enable the decision-making process to be carried out within a few weeks using the knowledge available to a typical process engineer. Its essence is to recommend solutions to the problem of tactical capacity planning, considering production and economic issues. The article attempts to respond to this challenge by proposing a tactical capacity planning method verified by a case study of a medium-sized roof window company. The essence of the suggested approach is to combine discrete event simulation (DES) with throughput accounting (TA). DES allows analysing the current production situation and aids to design change scenarios. TA is used to diagnose financial effects based on data from simulation experiments and estimated costs of implementing changes.

The proposed approach to the capacity planning issue is assigned primarily to small and medium-sized enterprises (SMEs). Limited time and competencies inhibit them to use methods based on mathematical programming or artificial intelligence, which are discussed in the next chapter of the article. It is also discussed in it the new opportunities for SMEs give by DES software development. Chapter three provides a detailed description of the problem, research methods, and the starting state of the analyzed manufacturing system. Chapter four shows the step-by-step process of planning of tactical capacity according to the proposed method. Chapter five discusses the significance of the results got in a broader context relating to operational management. Chapter six outlines the advantages of the proposed method and points to limitations in its application.

2. Theoretical background

2.1 Overview of tactical capacity planning methods

Capacity can be defined as the total productive capability of all the utilized productive resources including workforce and machinery [9]. Capacity management aims to ensure that a manufacturer has the 'right' capacity to act within a complex structure and how best to 'utilise' their internal capabilities [4]. The capacity decisions can be made in all decision-making hierarchies: strategic, tactical and operational [9]. The scope of tactical capacity planning decisions includes [10]:

- considering the mid-term demand, to decide whether the capacity of the existing factories should be expanded by buying new auxiliary tools (purchase or replace bottleneck machines; purchase or replace auxiliary tools),
- considering the mid-term demand, to decide the best portfolio of products (decide the best portfolio of products).

The analysis in this article is limited to the first scope of decisions oriented towards increasing production capacity through changes in the production process.

Many of the solutions to the capacity planning problem presented in the literature are based on the Integer Linear Programming (LP) or Mixed Integer Programming (MIP) models, [11, 12]. They work very well in large, automated companies, as well as with production allocation problems in various production facilities [13], or the combination of strategic and tactical decisions in the design of the distribution network including storage capacity, [14]. The problem is the adequacy of linear models, which is often not satisfactory. Complex systems may not reflect the decision-making situation well enough when uncertainties need to be taken into account. In such cases, models based on stochastic programming [15] are used. The literature reports good results in the use of such an approach [16], but the high complexity and difficulty of using them leads to the search for metaheuristic algorithms combining deterministic and stochastic approaches [17].

A large group of tactical capacity planning solutions is based on combining simulation and optimization techniques. Examples include the use of artificial neural networks in connection with genetic neural networks [18], or the link between genetic algorithms and Monte Carlo simulation [19]. However, the use of such solutions requires advanced knowledge and time, which is why the publications presenting them refer to industries dominated by large companies. It should also be noted that hasty imposition in the model of certain assumptions may lead to a situation where the real problem of the company is not solved.

The latest capacity planning solutions take advantage of the opportunities offered by Industry 4.0 application. The capacity planning model for prefabricated housebuilding elements uses real-time data entry to improve assembly line performance planning [20]. This approach develops short-term capacity equalization in make-to-order production systems [21]. Small and medium-sized enterprises, however, generally do not have sufficient resources to collect large amounts of data or do not have the skills to process it [22]. The solution proposed for them is a computer simulation, increasingly recommended in publications based on practical problems presented on the basis of a case study of enterprises [23].

2.2 Discrete event simulation in manufacturing capacity development

Simulations represent imitation of some process in the real world. Process simulation often requires generating a model that will represent the key characteristics of the selected system or process and will present its behaviour [24].

Discrete event simulation (DES) is an instrument for analysing dynamic processes taking into account stochastic parameters and uncertain events [25]. Simulation modelling aims to qualitatively or quantitatively support decisions by building a model of a real system and experimenting with that model [26]. The increasing importance of simulation methods in solving complex production problems is demonstrated by the growing number of studies in this field [27]. Among them, methods based on discrete event simulation [28] occupy an important place. They support the resolution of many different capacity planning problems, such as:

- configuration of flexible modular mounting systems [29];
- matching capacity to uncertain demand for hospitals and other healthcare providers [30];
- analysis of employee productivity in a production process with high variability due to low level of automation [31];
- decision-making tool in the complex production configuration of the automotive assembly line [32].

DES modelling uses the software of various types. Commercial systems such as Tecnomatix Plant Simulation, ARENA, Enterprise Dynamics and FlexSim are the most popular among industrial engineers. The first was used in the presented studies because it provides advanced options for each component of the model to accurately match it to the actual production system. Also not without significance is the extensive help in the form of online guides and training offered by the manufacturer and the community using this software. Open source systems such as Salabim, JaamSim, whose usability is compared to commercial systems [33] are also becoming increasingly popular.

The greatest benefit of using DES in capacity planning is the ability to represent even complex processes without losing the accuracy of the results or significantly increasing the calculation time. Conducting experiments using DES provides an understanding of the system's behaviour before it is built, the detection of unexpected incompatibilities, the ability to investigate various applications of case scenarios [34]. However, it should be remembered that simulation models only allow you to predict the outcome of different scenarios. Thus, the "best" solution is chosen from the scenarios analysed, not from all possible, in contradistinction to optimization models, where the "best of all" is expected. The use of computer simulation, therefore, does not guarantee an optimal solution to the problem of capacity planning.

3. Problem statement and used methodology

3.1 Problem statement

A research gap is identified regarding the use of decision support tools to solve tactical planning problems in real-world production conditions [35]. The article considers the problem of capacity planning in the company manufacturing roof windows. It is one of the manufacturing plants for a well-known brand in Europe. Of the 150 people employed, 122 serve production and warehouses, and the rest are technical and administrative staff. The company's current production capacity is 800 pcs/business day when working on two shifts. The plant received from the headquarters (from the window distributor) a question about the possibility of increasing production for half a year to 1100 pcs/day, a decision must be made within 1 month, and possible changes made within 3 months. At the moment, the company makes such decisions based on the opinions of internal specialists. The analyses are limited to a specific brainstorming session in which 2-3 variants of change are considered. This approach, in the opinion of managers, is insufficient. They would like the decision-making process for capacity planning to be based on a structured decision-making process supported by reports containing reasonable estimates of the effects of change. The analysis time offered by consulting companies is not adequate to the requirements of tactical planning. In addition, the company has experienced and committed employees involved in the development of its competencies.

3.2 Research objectives

The practical purpose of the research was to find an answer to the question of how the tactical capacity planning decision-making process should proceed so that it could be used by the window manufacturer. In a broader context, the research aims to propose a capacity planning methodology that is useful for the resources available to small and medium-sized enterprises. This includes providing theoretical guidance on how to approach the design of organisational changes and to assess their production and financial performance in the real world. Combining them into an orderly methodology is intended to promote rationalisation of the decision-making process, taking into account the constraints typical of small and medium-sized enterprises. Qualitative studies that answer the 'how' question require an understanding of the conditions for implementing the methodology and should therefore be supported by case study methods [36].

3.3 Methodology

Most of these scientific publications in the field of capacity management seek to obtain practical results by confronting existing theories with problems occurring in industrial reality. In the field of operations management, case research is seen as one of the most powerful methods [37]. The case study method is very often used in the study of production systems because it allows for a comprehensive and broad view of complex issues in highly diverse environments [38]. The research presented in the article leads to the development of solutions that are valuable from a practical point of view while generating theoretical knowledge. Efforts were made to present in detail the data collected during the analysis of the current situation of the production system. They were collected by:

- observation of the production system by the investigator, which made it possible to identify the elements of the system and understand the connections between them,
- testing of production documents, including, in particular, reports from the SAP ERP system,
- measurements made in the production system for the preparation of its computer model,
- semi-structured interviews on production system constraints and unstructured interviews relating to the company's capacity planning capabilities and needs,
- simulation experiments leading to verification of the developed model.

The data collected in this way was used to:

- development of a computer model illustrating the current state of the production system,
- formulate assumptions for change scenarios,
- develop a concept for evaluating change scenarios.

In the second phase of the study, the following research techniques were used to develop and analyse change scenarios:

- participant observation when designing change scenarios,
- experiments based on computer simulation to determine the production effects of a given change scenario,
- structured interview on estimating the financial parameters of the change scenario.

The aim of the second phase was to characterise the scenarios of change in order to recommend one of them or to reject the proposal to increase production capacity.

3.4 Model of current state of roof window plant

The case study involved a roof window manufacturer that is part of a global network of sales, production and services. The article only analyses the assembly line bypassing the process of painting components and material handling, because their flexibility is much higher than the assembly process and allows a much larger volume of production. There are 20 workstations on the line. The daily production plan is generated by the SAP ERP system, which determines the order in which orders are run. The plant works two shifts five days a week. The current demand volume is 800 units/day. In the system, you can distinguish 5 sections presented in Fig.1. The description of which is given in Table 1.

The upper part of Fig. 1 shows the layout of the entire assembly line, which distinguishes sections such as interior wing assembly line (In_sash), external wing assembly line (out_sash), installation station of both window wings (Sash_assembly), window assembly section (pane_assembly), packaging section (Packaging). The lower part of Fig. 1 presents the positions in the different sections of the production line. The model was made using Tecnomatix Plant Simulations by Siemens software.

In the section where the outer wing is joined (sash_assembly) and in the section where the glass is inserted, there is a large share of manual work. The triangular distribution was used for its modelling. It is defined by three parameters (c, a, b), where the most probable value is c, the minimum value is a and the maximum value is b. In Tecnomatix Plant Simulation triangular distribution is called triangle distribution. Table 1 describes the object parameters in each section of the model and describes them. This data allows you to recreate the model in any DES environment.

Based on production reports from SAP – ERP, it is assumed that the frequency of overarching best aggregates the batch size of 25 units. In practice, the size of the production batches varies and is a multiple of 5 pcs.

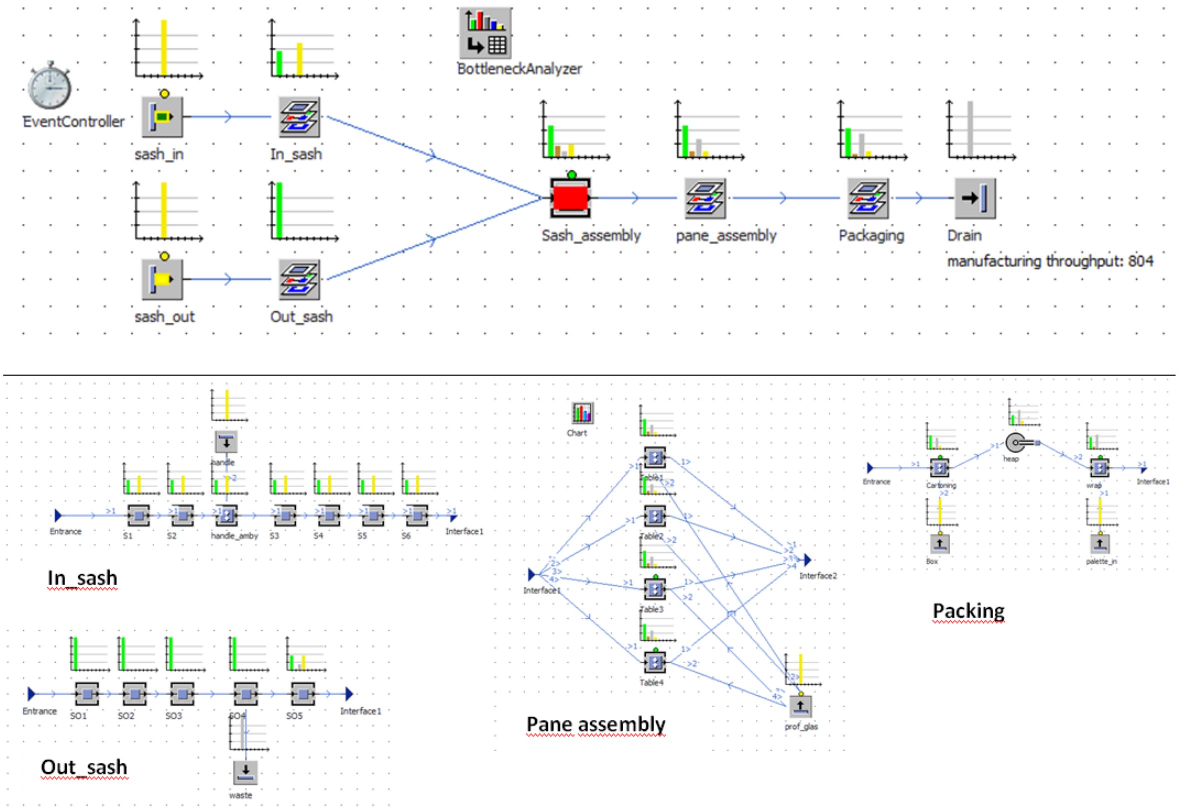


Fig. 1 Assembly line model of the roof window manufacturer made in Technomatix Plant Simulation system

Table 1 Description of the individual positions of the modelled production process

Section description	Model marking	Section characteristics
Installation of internal wing elements	*In_sash* One worker on each station	A string of 7 mounting tables bound by roller relays. Processing time for each position 30 s.
Installation of external wing elements	*Out_sash* One worker on each station	A string of 7 mounting tables bound by roller relays. Processing time for each position 30 s. In position SO4 1 % waste.
Installation of the outer and inner wings	*Sash_assembly* Two cooperated workers	Processing time: triangle (0:40, 0:20, 1:00) Set-up time: triangle (5:30, 4:15, 6:00) Set-up after 25 parts before next part
Installation of the glass	*Pane_assembly* One worker on each station	4 independent stations equipped with slatted glass mounting tables and aluminium profile mounting machines. Processing time: triangle (2:10, 1:30, 4:00) Set-up time: triangle (8:00, 6:00, 15:00) Set-up after 25 parts before next part.
Packing	Packing One worker on each station cartoning	Equipment assembly: cartoning machine, robot for stacking windows into packages of 5 pieces (heap), foiling machine and preparation of packages for transport (wrap). Processing time: 35 s Set-up time: 1 min Set-up after 25 parts before next part
	heap	Loading time 5 s Unloading time 10 s
	wrap	Processing time: 30 s Set-up time: triangle 20 s Set-up after 5 parts before next part.

4. Results and discussion: A case study for designing and analysing changes in the production system

4.1 Proposed approach to designing change scenarios in manufacturing system

The proposed methodology for capacity planning is largely based on the Theory of Constraints logic. If demand exceeds the company's capacity in terms of production volume and is of a physical nature, it identifies a resource constraint [39]. The process of improving the production system should begin with such a limitation called a bottleneck [40]. The next step in improving the production system is to determine how to exploit the bottleneck. Typically, there are more than a dozen potential improvement scenarios focusing on bottlenecks [38]. Their development takes time and is determined by the experience and creativity of the management team.
Because of the limited time, it is proposed to design change scenarios taking into account the decreasing likelihood of their success. It is assumed that it is due to difficulties relating to the organisational effort of the staff and the risks associated with the novelty of the changes, which affect the uncertainty about the expected results and the error of financial estimates. The following order in which change scenarios are formulated is proposed:

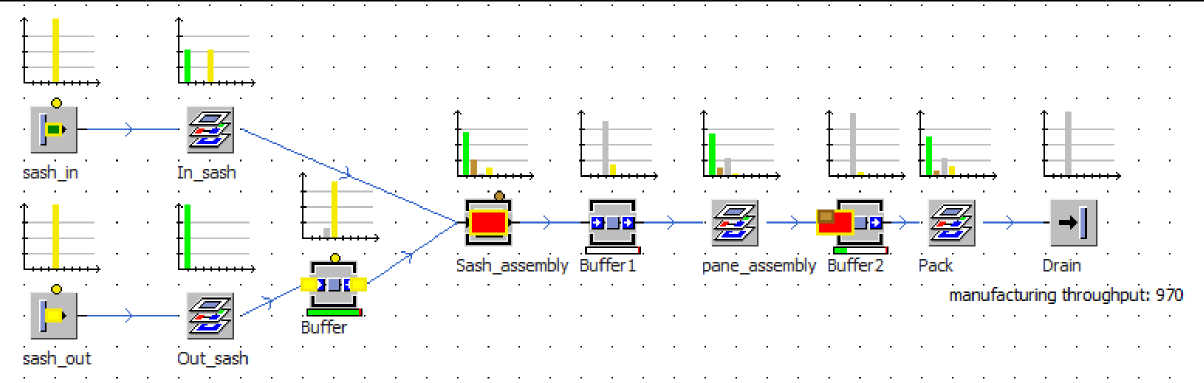
- Balance the production flow to a consistent rate imposed by the bottleneck.
- Elimination of losses on bottleneck according to Lean Production methods.
- Introduction of cooperation and outsourcing in the production process.
- Introducing incremental and radical innovations in the production process.

4.2 Assessment of production parameters of changes scenarios

Bottleneck-focused improvements are typically strongly linked to other components of the production system, so the capacity increase of the entire production system should be estimated as a basis for their implementation. The use of DES in the proposed methodology is not limited to analysing the current state of the manufacturing system but is extended to include an analysis of future conditions on the basis of "what if" experiments. Table 2 presents a summary of the analysed scenarios of change with scenario details, where t_i means processing time and t_{su} means set-up time. In addition, the scenario resulting from the frequent belief that in the event of a shortage of capacity, it is sufficient to install a new manufacturing unit, a scenario marked with the letter B, which involves the duplication of the mounting position of the inner wing and the outer roof window.

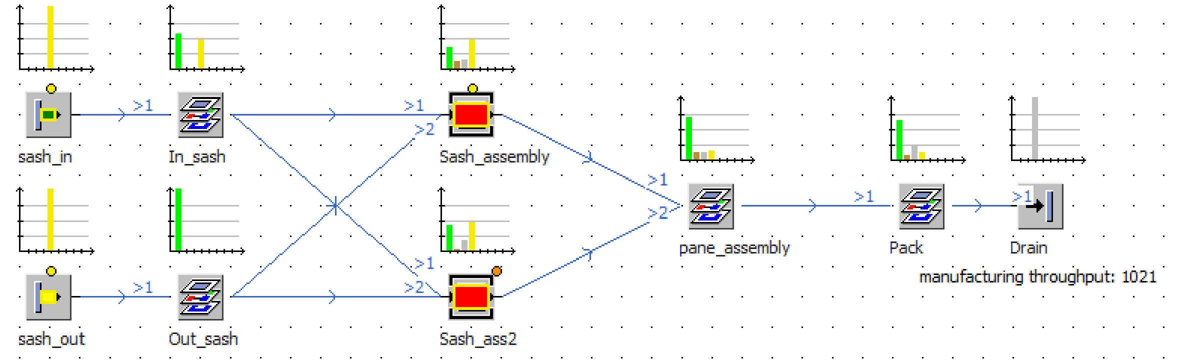
Table 2 Summary of analysed scenarios of changes in the production system

A: Put the three buffers into the manufacturing system for blockade limitation on the stations: *sash_assembly* and *pane_assembly*. Buffers are on the tables with roller conveyers.
Scenario details: Buffer capacity = 2, Buffer1 capacity = 4, Buffer2 capacity = 4
Total throughput of the manufacturing system after changes: 970 roof window

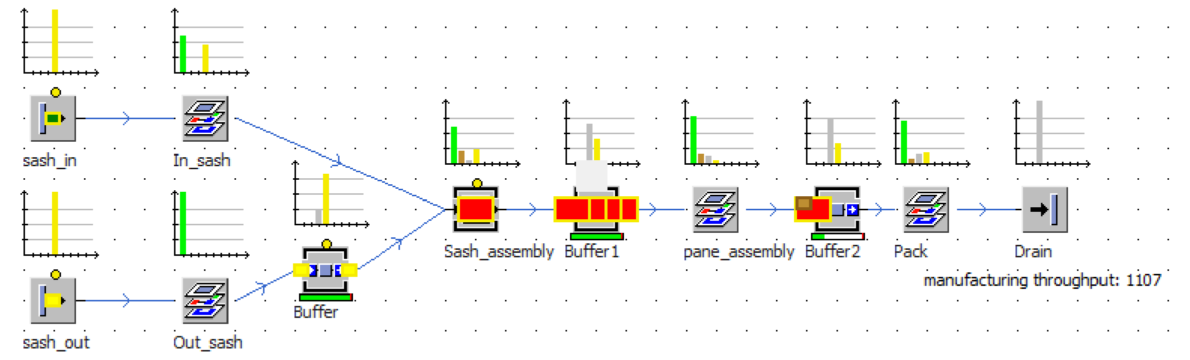


B: Duplication the *sash_assembly* station.
Scenario details: Interposing stadion: Sash_ass2; t_i -> (0:40, 0:20, 1:00)F; t_{su} -> (5:30, 4:15, 6:00)
Total throughput of the manufacturing system after changes: 1021 roof window

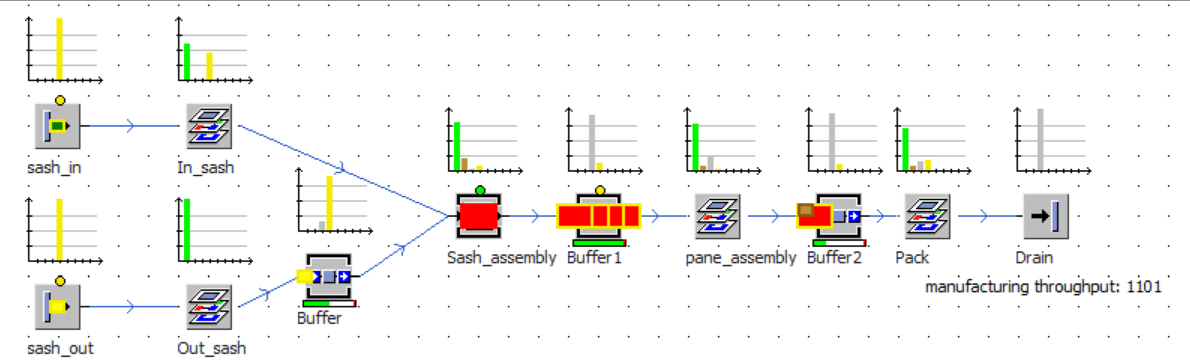
Table 2 Summary of analysed scenarios of changes in the production system (continuation)



A+C: Case A combined with introducing the standardization on the station *sash_assembly* for shortening processing and setup time.
Scenario details: Buffer capacity = 2, Buffer1 capacity = 4, Buffer2 capacity = 4, Sash_assembly: t_i -> (0:30, 0:20, 0:40), t_{su} -> (4:00, 3:00, 4:30)
Total throughput of the manufacturing system after changes: 1107 roof window

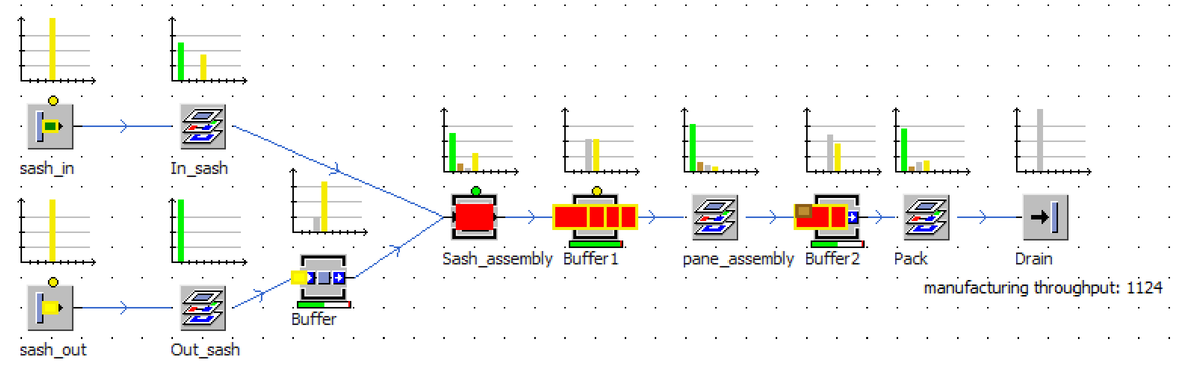


A+D: Case A combined with additional worker on pane_assembly, who help in set-up. The scenario assumes panel computer installation, which shows setups ahead.
Scenario details: Buffer capacity 2, Buffer1 capacity = 4, Buffer2 capacity = 4, Sash_assembly: t_{su} -> (4:00, 3:00, 4:30), Table1-Table4: t_{su} -> (4:00, 3:00, 7:00)
Total throughput of the manufacturing system after changes: 1101 roof window

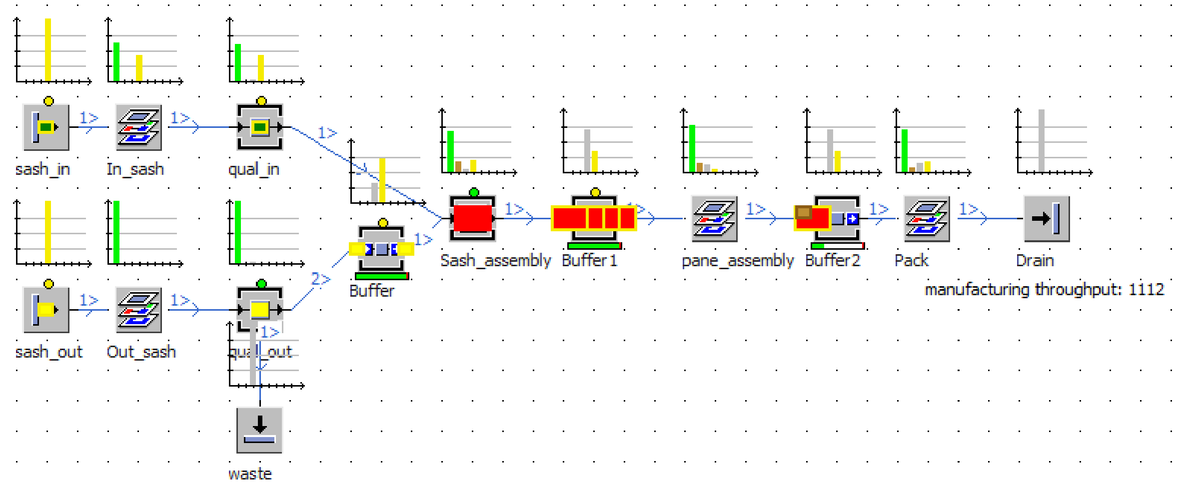


A+E: Case A combined with supporting operations on sash_assembly station by cobot.
Scenario details: Buffer capacity = 2, Buffer1 capacity = 4, Buffer2 capacity = 4, Sash_assembly: t_i -> (0:30, 0:20, 0:40) t_{su} -> (2:30, 1:15, 3:00)
Total throughput of the manufacturing system after changes: 1124 roof window

Table 2 Summary of analysed scenarios of changes in the production system (continuation)



A+F: Case A combined with vision system(worker free) for sash quality (scratch, dimension, angles, colour) control.
Scenario details: Buffer capacity 2, Buffer1 capacity 4, Buffer2 capacity 4, Sash_assembly: t_i -> (0:30, 0:20, 0:50) t_{su} -> (3:30, 2:15, 4:00), Station qual_in: t_i -> 0:30 Station qual_out: t_i -> 0:30
Total throughput of the manufacturing system after changes: 1112 roof window



4.3 Estimating the financial impact of production system change scenarios

It is not sufficient to limit the decision-making process solely to assessing the production effects of change options. Recommending the selection of one of them requires at least taking into account the financial parameters of the scenario. In the methodology being developed, it is proposed to use throughput accounting (TA) to estimate the financial impact of the options for change being analysed. TA is a simplified management accounting method that provides managers with support in making decisions aimed at increasing the profitability of the company. The undoubted advantage of TA is simplicity in assessing the investment made through practical and simple measures that take into account the impact of the changes on the whole functioning of the company. Those that have been used in the proposed methodology are presented in Table 3.

Table 3 TA meters used to estimate the cost-effectiveness of change scenarios

Throughput = Sales Revenue – Totally Variable Cost T = SR – TVC	Throughput (T) is net sales (SR) less totally variable cost (TVC), generally the cost of the raw materials, transportation charges, outsourced processing, commissions deducted for sales. *Labour and energy in manufacturing organizations are usually not tied to units produced.
Net profit = Throughput – Operating Expense NP = T – OE	Operating Expenses: Operating expenses (OE) refer to other cash outflows needed for creating the throughputs like payments made for salaries and benefits of employees, maintenance, rental expenses, lease expenses, taxes and license fees, cost of utilities, etc.
Return on Investment = Net Profit / Investment ROI = NP / I	Investments (I) represent funds tied up in physical assets such as machinery and equipment, land and buildings, product inventory, etc. In the TOC these all referred to as inventory

Based on: [40]

TA is designed as a direct cost approach and as such supports in particular short- and medium-term production decisions [41]. Although it does not provide such precise answers as traditional accounting, it is sufficiently accurate for tactical investment decisions. Table 4 compiles estimates of financial parameters for each production system change scenario. It is assumed that the throughput/unit is 5 €, and during the year the company work for 250 days. It should be noted that the investment costs have been estimated and are all the more precise the less innovative the solution proposed in the scenario.

The very high ROI results show that the company should decide to make investments to increase production capacity to 1100 units per day. The most cost-effective option was the introduction of three buffers to mitigate the effects of variable technological operations times and standardisation at the installation site of the inner wing and the outer window (scenario C). This solution provides for the introduction of additional bonuses for employees in this position of € 150/day, but the assumed effects in the form of a capacity increase of € 1480/day justify this expenditure. Such an increase in earnings (about 40 %) should ensure that workers with key skills are maintained over a longer period of time. The danger was that the quality would be reduced, so when implementing this solution, the company decided to make a 4h shift for each employee in this position and move it to a neighbouring position (pane assembly).

Table 4 Financial parameters of the analysed changes in the production system

	A	B	A + C	A + D	A + E	A + F
Δunits/day	166	217	296	296	296	296
ΔT/day (€)	830	1085	1480	1480	1480	1480
ΔOE (€/day)	12.5	492	160.5	153.5	40.9	43.5
Δdepreciation €/day	10.5	8	10.5	18.5	32.9	22.5
Δsalary €/day	0	480	150	120	0	0
Δmaintenance €/day	2	2	0	10	4	15
Δenergy €/day	0	2	0	5	4	6
ΔNP (€/day)	817.5	593	1319.5	1326.5	1439.1	1436.5
ΔNP (€/year)	204375	148250	329875	331625	359775	359125
ΔI (€)	2625	2000	3125	4625	8225	5625
ROI = ΔNP/ΔI	77.86	74,13	**105.56**	71.70	43.74	63.84

4.4 Discussion

The study presented was based on a case study of a medium-sized production company. It showed the use of the proposed capacity planning methodology to answer the question of which of the available scenarios for changes that adapts capacity to demand should be implemented by the company. Analyses leading to the selection of the change plan option were carried out within a month. It should be noted that the development of the computer model of the system itself (together with the collection of data) took 2 weeks. If such a model were already available at the time of the question of the possibility of changing capacity, the response time could be reduced, or more change scenarios could be designed.

A key advantage of computer simulation approaches is the ability to deal with complex stochastic problems without the need for advanced mathematical skills [42]. The article shows how DES can easily support mid-level managers in managing the company's operations. Although it is not a precise tool, it is accurate enough to allow for rational decision-making in the case of medium-term investments. DES systems are constantly being improved. In addition, the growing offer of textbooks, tutorials, and discussion forums makes it possible to acquire competencies sufficient to model a production system in a specific enterprise relatively quickly (about a month). This task is also facilitated by access to production documentation from ERP class systems. Thus, the limitations of the availability of computer simulation tools for SMEs due to knowledge, cost, data availability and development time (limited by knowledge, cost, data availability and development time) [43] are systematically losing their importance.

The methodology proposes to design change scenarios in a predetermined sequence. Thus, plans are created with increasing complexity, which allows, among other things, for combining scenar-ios of change. This was the case, where further capacity expansion was based on the implementation of scenario A and expanding it with new elements. The initial simulations already concluded that buffers are the solution to the volatility resulting from short-term stochastic irregularities [44]. Their introduction involves the dilemma of increasing work-in-process (WIP)to fully utilize resources.

In the proposed methodology, the investment profitability analysis is evaluated based on the ROI determined according to the TA principles. The assumptions of the cost analysis TA method used proved simple to understand for the engineers implementing the project. However, this approach does not account for the uncertainty with which the costs of innovative solutions are estimated. In the case study, this applies to the scenarios labelled A+E and A+F.

The specificity of the object of the investigation lies in the fact that the company does not forecast demand itself, but rather adjusts its production capacity to the limits imposed by the central distributor. This is quite a typical situation in the case of Polish companies, which often carry out production for European brands.

The proposed methodology is adequate for enterprises implementing complex discrete production. The effectiveness of the method developed has been demonstrated by applying it in a real-life production process. The results show that using the associated DES and TA tools allow managers to successfully design scenarios of changes in the production system, analyze them and assess their financial impact. The rationalization of tactical capacity planning enables a quick response to emerging opportunities to increase revenues related to changes in demand or reduce costs by implementing new technological solutions.

It is a challenging skill that constitutes a competitive advantage in various SME enterprises with complex production systems. The use of DES software enables the design and analysis of change scenarios taking into account technical constraints. It allows the extrapolation of production capacity, providing the basis for economic assessments. The proposed approach is the foundation for planning the development of a production system that have regard both technical constraints and profit-making aspects.

5. Conclusions and future research

Digital models and simulation experiments make it possible to bring a lot of relevant information into the capacity planning process that is not otherwise possible in practice. The importance of DES in tactical capacity planning stems from the fact that different decision options for organising the process can be tested before they are implemented. The concept of combining DES and TA-based financial analysis is related to the need for decision support that:

- must be carried out in a relatively short and limited period of time,
- has little recourse to external consultancy, (e.g. for financial analysis),
- is not routine and may not follow a set pattern, such as the entry of overtime,
- requires a good understanding of the production situation.

Among other things, tactical capacity planning decisions are made under such conditions. The proposed methodology enables a small team of engineers to find answers to the questions: what will happen if certain changes are made [1], where their weaknesses are [2], how individual change scenarios will affect the capacity of the overall system [3]. However, it must be emphasized that the proposed methodology does not lead to an optimal decision. Nor does it guarantee to find solutions that are workable. It requires employees involved in the decision-making process to be creative in designing change scenarios and skilled in operating the computer system.

The need to support decision-making processes in terms of tactical changes in production systems is not only due to the need to match capacity to demand. It is also related to the analysis of the effects of implementing solutions that fit into the Industry 4.0 concept. Many SMEs are already facing a dilemma regarding this type of investment. The proposed methodology makes it possible to estimate the impact of changes on system capacity and seems adequate to support this type of decision as well. However, this requires verification of not only the manufacturing system but the entire production system. Research on the use of the methodology in evaluating scenarios for introducing elements of the Industry 4.0 concept in a medium-sized manufacturing company is currently being conducted. Furthermore, the subject of the research is also the introduction of multi-criteria decision - making methods for the evaluation of change scenarios in terms of sustainable development criteria resulting from the corporate strategy.

References

[1] Seyedan, M., Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities, *Journal of Big Data*, Vol. 7, Article No. 53, doi: 10.1186/s40537-020-00329-2.

[2] Kampa, A., Gołda, G., Paprocka, I. (2017). Discrete event simulation method as a tool for improvement of manufacturing systems, *Computers,* Vol. 6, No. 1, Article No. 10, doi: 10.3390/computers6010010.

[3] Patalas-Maliszewska, J., Topczak, M. (2021). A new management approach based on additive manufacturing technologies and Industry 4.0 requirements, *Advances in Production Engineering & Management*, Vol. 16, No. 1, 125-135, doi: 10.14743/apem2021.1.389.

[4] Sabet, E., Yazdani, B., Kian, R., Galanakis, K. (2020). A strategic and global manufacturing capacity management optimisation model: A scenario-based multi-stage stochastic programming approach, *Omega,* Vol. 93, Article No. 102026, doi: 10.1016/j.omega.2019.01.004.

[5] Jacobs, F.R., Chase, R.B. (2008). *Operations and supply management: The core,* McGraw Hill/Irwin, New York, USA.

[6] Zhang, R.-Q. (2007). Research on capacity planning under stochastic production and uncertain demand, *Systems Engineering – Theory & Practice*, Vol. 27, No. 1, 51-59, doi: 10.1016/S1874-8651(08)60006-X.

[7] Gallego-García, S., Reschke, J., García-García, M. (2019). Design and simulation of a capacity management model using a digital twin approach based on the viable system model: Case study of an automotive plant, *Applied Sciences*, Vol. 9, No. 24, Article No. 5567, doi: 10.3390/app9245567.

[8] Sawik, T. (2009). Multi-objective due-date setting in a make-to-order environment, *International Journal of Production Research*, Vol. 47, No. 22, 6205-6231, doi: 10.1080/00207540902810585.

[9] Alp, O., Tan, T. (2008). Tactical capacity management under capacity flexibility, *IIE Transactions*, Vol. 40, No. 3, 221-237, doi: 10.1080/07408170701488052.

[10] Chen, Y.-Y., Chen, T.-L., Liou, C.-D. (2012). Medium-term multi-plant capacity planning problems considering auxiliary tools for the semiconductor foundry, *The International Journal of Advanced Manufacturing Technology*, Vol. 64, No. 9-12, 1213-1230, doi: 10.1007/s00170-012-4080-9.

[11] Tavaghof-Gigloo, D., Minner, S., Silbermayr, L. (2016). Mixed integer linear programming formulation for flexibility instruments in capacity planning problems, *Computers & Industrial Engineering*, Vol. 97, 101-110, doi: 10.1016/j.cie.2016.04.013.

[12] Zschorn, L., Müller, S., Ivanov, D. (2017). Capacity planning on key work stations in a hybrid MTO-ETO production system: A case-study on Siemens AG, *International Journal of Inventory Research*, Vol. 4, No. 2-3, 214-232, doi: 10.1504/ijir.2017.088743.

[13] Fleischmann, B., Ferber, S., Henrich, P. (2006). Strategic planning of BMW's global production network, *INFORMS Journal on Applied Analytics,* Vol. 36, No. 3, 194-208, doi: 10.1287/inte.1050.0187.

[14] Yu, V.F., Normasari, N.M E., Luong, H.T. (2015). Integrated location-production-distribution planning in a multi-products supply chain network design model, *Mathematical Problems in Engineering*, Vol. 2015, Article ID 473172, doi: 10.1155/2015/473172.

[15] Lin, J.T., Chen, T.-L., Chu, H.-C. (2014). A stochastic dynamic programming approach for multi-site capacity planning in TFT-LCD manufacturing under demand uncertainty, *International Journal of Production Economics*, Vol. 148, 21-36, doi: 10.1016/j.ijpe.2013.11.003.

[16] Wang, K.-J., Nguyen, P.H. (2017). Capacity planning with technology replacement by stochastic dynamic programming, *European Journal of Operational Research*, Vol. 260, No. 2, 739-750, doi: 10.1016/j.ejor.2016.12.046.

[17] Geng, N., Jiang, Z. (2009). A review on strategic capacity planning for the semiconductor manufacturing industry, *International Journal of Production Research*, Vol. 47, No.13, 3639-3655, doi: 10.1080/00207540701871051.

[18] Teerasoponpong, S., Sopadang, A. (2021). A simulation-optimization approach for adaptive manufacturing capacity planning in small and medium-sized enterprises, *Expert Systems with Applications*, Vol. 168, Article No. 114451, doi: 10.1016/j.eswa.2020.114451.

[19] Renna, P. (2016). A decision investment model to design manufacturing systems based on a genetic algorithm and Monte-Carlo simulation, *International Journal of Computer Integrated Manufacturing*, Vol. 30, No. 6, 590-605, doi: 10.1080/0951192x.2016.1187299.

[20] Huka, M.A., Grenzfurtner, W., Zauner, B., Gronalt, M. (2021). Capacity planning of a mixed-model assembly line for prefabricated housebuilding elements, *Procedia Computer Science*, Vol. 180, 706-713, doi: 10.1016/j.procs.2021.01.293.

[21] Zhang, B.-X., Song, J., Yu, X.-L. (2012). Research about short-term production capacity decision-making, *Physics Procedia*, Vol. 24, Part B, 1207-1213, doi: 10.1016/j.phpro.2012.02.180.

[22] Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., Reis, M.S. (2016). How can SMEs benefit from big data? Challenges and a path forward, *Quality and Reliability Engineering International*, Vol. 32, No. 6, 2151-2164, doi: 10.1002/qre.2008.
[23] Straka, M., Khouri, S., Lenort, R., Besta, P. (2020). Improvement of logistics in manufacturing system by the use of simulation modelling: A real industrial case study, *Advances in Production Engineering & Management*, Vol. 15, No. 1, 18-30, doi: 10.14743/apem2020.1.346.
[24] Kotevski, Ž., Jovanoski, B., Minovski, R. (2015). Simulation model for improved production planning and control through quality, cycle time and batch size management, *Journal of Engineering Management and Competitiveness*, Vol. 5, No. 1, 40-45, doi: 10.5937/jemc1501040k.
[25] Turk, M., Pipan, M., Šimic, M., Herakovič, 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.
[26] Bangsow, S. (2010). *Manufacturing simulation with plant simulation and simtalk: Usage and programming with examples and solutions,* Springer, Berlin, Germany, doi: 10.1007/978-3-642-05074-9.
[27] Zhang, L., Zhou, L., Ren, L., Laili, Y. (2019). Modeling and simulation in intelligent manufacturing, *Computers in Industry*, Vol. 112, Article No. 103123, doi: 10.1016/j.compind.2019.08.004.
[28] Gansterer, M., Almeder, C., Hartl, R.F. (2014). Simulation-based optimization methods for setting production planning parameters, *International Journal of Production Economics*, Vol. 151, 206-213, doi: 10.1016/j.ijpe.2013.10.016.
[29] Gyulai, D., Monostori, L. (2017). Capacity management of modular assembly systems, *Journal of Manufacturing Systems*, Vol. 43, Part 1, 88-99, doi: 10.1016/j.jmsy.2017.02.008.
[30] Bae, K.-H., Jones, M., Evans, G., Antimisiaris, D. (2019). Simulation modelling of patient flow and capacity planning for regional long-term care needs: A case study, *Health Systems,* Vol. 8, No. 1, 1-16, doi: 10.1080/20476965.2017.1405873.
[31] Antonelli, D., Litwin, P., Stadnicka, D. (2018). Multiple system dynamics and discrete event simulation for manufacturing system performance evaluation, *Procedia CIRP*, Vol. 78, 178-183, doi: 10.1016/j.procir.2018.08.312.
[32] Sarda, A., Digalwar, A.K. (2018). Performance analysis of vehicle assembly line using discrete event simulation modelling, *International Journal of Business Excellence*, Vol. 14, No. 2, 240-255, doi: 10.1504/ijbex.2018.10009766.
[33] Lang, S., Reggelin, T., Müller, M., Nahhas, A. (2021). Open-source discrete-event simulation software for applications in production and logistics: An alternative to commercial tools?, *Procedia Computer Science*, Vol. 180, 978-987, doi: 10.1016/j.procs.2021.01.349.
[34] Istokovic, D., Perinic, M., Dobovicek, S., Bazina, T. (2019). Simulation framework for determining the order and size of the product batches in the flow shop: A case study, *Advances in Production Engineering & Management*, Vol. 14, No. 2, 166-176, doi: 10.14743/apem2019.2.319.
[35] Carvalho, A.N., Scavarda, L.F., Oliveira, F. (2017). An optimisation approach for capacity planning: Modelling insights and empirical findings from a tactical perspective, *Production*, Vol. 27, Article No. e20170014, doi: 10.1590/0103-6513.001417.
[36] Yin, R.K. (2003). *Case study research: Design and methods,* Third Edition, Sage PublicationsThousand Oaks, USA.
[37] Voss, C., Tsikriktsis, N., Frohlich, M., (2002). Case research in operations management, *International Journal of Operations & Production Management*, Vol. 22, No. 2, 195-219, doi: 10.1108/01443570210414329.
[38] Urban, W. (2019). TOC implementation in a medium-scale manufacturing system with diverse product rooting, *Production & Manufacturing Research*, Vol. 7, No. 1, 178-194, doi: 10.1080/21693277.2019.1616002.
[39] Baxendale, S.J., Raju, P.S. (2004). Using ABC to enhance throughput accounting: A strategic perspective, *Cost Management*, Vol. 1-2, 31-38.
[40] Corbett, T. (1998). *Throughput accounting: TOC's management accounting system*, North River Press, Great Barrington, USA.
[41] Dražić Lutilsky, I., Liović, D., Marković, M. (2018). Throughput accounting: Profit-focused cost accounting method, In: *Proceedings of International Conference Interdisciplinary Management Research XIV*, Opatija, Croatia, 1382-1395.
[42] Figueira, G., Almada-Lobo, B. (2014). Hybrid simulation–optimization methods: A taxonomy and discussion, *Simulation Modelling Practice and Theory*, Vol. 46, 118-134, doi: 10.1016/j.simpat.2014.03.007.
[43] Aggogeri, F., Faglia, R., Mazzola, M., Merlo, A. (2015). Automating the simulation of SME processes through a discrete event parametric model, *International Journal of Engineering Business Management*, Vol. 7, 4 pages, doi: 10.5772/59958.
[44] Kim, S., Cox, J.F., & Mabin, V.J. (2010). An exploratory study of protective inventory in a re-entrant line with protective capacity, *International Journal of Production Research*, Vol. 48, No. 14, 4153-4178, doi: 10.1080/00207540902991666.

APEM journal
Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 348–358
https://doi.org/10.14743/apem2021.3.405
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Simulation-based optimization of coupled material–energy flow at ironmaking-steelmaking interface using One-Ladle Technique

Hu, Z.C.[a,b], Zheng, Z.[a,*], He, L.M.[c], Fan, J.P.[b], Li, F.[b]

[a]College of Materials Science and Engineering, Chongqing University, Chongqing, P.R. China
[b]CISDI Engineering Co., Ltd., Chongqing, P.R. China
[c]Chongqing College of Electronic Engineering, Chongqing, P.R. China

A B S T R A C T

The ironmaking-steelmaking interface of the steel manufacturing process involves the hot metal ladle circulation and the energy dissipation which are coupled processes with an interrelated but independent relation. Therefore, the synergistic operation of the material flow and the energy flow at the interface is momentous to the effective production of the ironmaking-steelmaking section. However, there is a lack of solutions to realize the synergy. Here, we presented a coupling simulation model for the material flow and energy flow of the ironmaking-steelmaking interface, based on the mathematical description of their operation behaviors, the operation and technical model of the production equipment and the temperature-decreasing model of the ladle. Further, the coupling simulation model was applied to a concrete ironmaking-steelmaking interface using the One-Ladle Technique. The coupling simulation model proved its performance in providing comprehensive decision-making supports and optimized production management strategies by achieving a solution that results in a decline of 10 °C in the average temperature drop of the hot metal and a reduction in the cost per tonne of steel by CNY 1.02.

A R T I C L E   I N F O

Keywords:
Metallurgy;
Ironmaking process;
Steelmaking process;
Ironmaking-steelmaking interface;
Coupled material-energy flow;
Discrete event simulation;
Optimization;
One-ladle technique

Corresponding author:
zhengzh@cqu.edu.cn
(Zheng, Z.)

Article history:
Received 8 August 2021
Revised 17 September 2021
Accepted 19 September 2021

1. Introduction

1.1 Problem background and description

The important ironmaking-steelmaking interface serves as the interconnection between the ironmaking plant and the steelmaking plant in the long-route steel manufacturing process. This section involves three processes, i.e. blast furnace (BF) for ironmaking, Kambara Reactor (KR) for hot metal pretreatment and basic oxygen furnace (BOF) for steelmaking. The transportation of the hot metal is the main logistics process in this section. The "One-Ladle Technique" of the ironmaking-steelmaking interface is a new technology for metallurgical process and transportation optimization emerged in the past decade [1]. The specific process is shown in Fig. 1: The empty ladles are pushed, usually, by a locomotive to the BF to charge the hot metal, after which the empty ladles become full ones. The full ladles are then transported to the buffer zone before the KR process. A crane lifts a ladle and transports it to the KR station for processing. Next, the ladle is again lifted by a crane to the BOF, into which the hot metal is poured. The full ladle becomes empty and then is directly put into the buffer zone by the same crane. Finally, the ladle is again transported by a locomotive to the BF and another circulation begins.

The ladle circulating rate affects the temperature drop of the hot metal and the ladle itself. In the circulation process, the material flow and the energy flow show an interweaved and interactive relationship. The multi-flow coupling phenomenon is the inherent characteristics that cannot be neglected when analyzing and optimizing the ironmaking-steelmaking interface [2]. The "One-Ladle Technique" puts forward higher requirements for ladle circulation, energy conversion, and time coordination in the production process. If too many ladles are running in the circulation, the charging waiting time increases, thereby increasing the temperature drop of the hot metal. However, the tapping temperature after the BOF process is constant, therefore the lower temperature of the hot metal poured into the BOF, the more heating agent is necessary to increase the temperature and longer processing time are required, which leads to a longer waiting time for the subsequent full ladle to be handled and a larger temperature drop. In summary, too many ladles will increase the waiting time, temperature drop, process processing time and auxiliary material cost. In turn, the lack of ladles will cause logistics interruption which affects the production. Therefore, the number of online ladles requires scientific decision-making.

The optimization analysis on the flow-coupling should focus on the optimization of logistics parameter configuration with the goal of reasonable cycle time, low production temperature drop, and low production cost. Therefore, aiming at the optimization of the ironmaking-steelmaking interface using the One-Ladle Technique, this paper develops a simulation model to conduct the analyses based on a real steelworks.

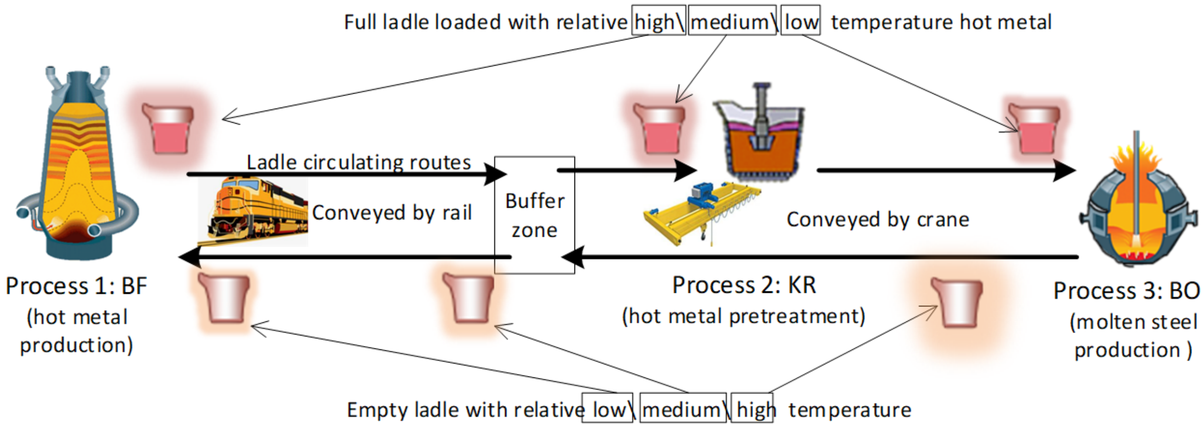


Fig. 1 Schematic of process flows at ironmaking-steelmaking interface

1.2 Previous work

Remarkable progress has been made in recent years in studying the ironmaking-steelmaking interface in terms of layout planning [3-8], connection mode of the interface [9-10], transportation scheduling [11-16], ladle circulating control [1, 2, 17-20], and hot metal temperature drop [21-23]. As regards layout planning, Fan *et al.* [3] analyzed and summarized the characteristics of three modes of transportation: railway, crane + transfer car and special roads at the ironmaking-steelmaking interface, and the relations with One-Ladle Technique and the general layout planning. Wiyaratn *et al.* [6] employed the SLP method in researching and evaluating the connection relationship between processes and compared the optimal schemes of the system under different layout plans. Concerning respect to the connections at the interface, Qiu *et al.* [7] analyzed the technological characteristics of the ironmaking-steelmaking interface in six typical process routes from the aspects of time, temperature, flow rate, production management, hot metal pretreatment effect, energy consumption, environmental pollution, etc. by using systematic scientific reduction theory and holism. As far as transportation scheduling is concerned, Tang *et al.*

[11, 12] studied the issues concerning the scheduling of arriving and departing torpedo ladle locomotives, and established a mixed integer programming model for scheduling of the hot metal transportation locomotives. And for ladle circulating control, Zhao *et al.* [1] adopted the multi-agent system to simulate the production logistics system of steel producers, and pointed out that the logistics efficiency should be improved by shortening refining cycle, increasing transportation speed, upgrading desulfurization equipment, etc. Xiao *et al.* [17] made analyses on possible optimization of production management schemes involving ladle preparation mode, number of online ladles at the ironmaking-steelmaking interface with One-Ladle Technique by material flow simulation. As for hot metal temperature drop, Du *et al.* [21] analyzed the heat dissipation mechanism of hot metal during hot metal receiving, transportation and pretreatment, etc. at the ironmaking-steelmaking interface from a heat transfer viewpoint, and established a mathematical model of hot metal temperature drop at the interface between hot metal ladle and charging ladle. Chen *et al.* [24] studied the fuel gas operation management practices for reheating furnace. Chen *et al.* [25] studied the effect of the production fluctuation on the process energy intensity in iron and steel industry, which revealed the relation of process energy intensity to the production operating rate and qualification rate.

In summary, the current research on the optimization of operation at the ironmaking-steelmaking interface focuses on optimization of the ladle circulation in the material flow or the temperature drop control in the energy flow but ignores the intrinsic rule of coupling control and synergy of material flows and energy flow during the operations at the interface. Operations at the ironmaking-steelmaking interface bear information on five dimensions: plan, time, equipment, operation, and temperature and composition of hot metal. As a single math or model method usually cannot take all the information of different dimensions into comprehensive consideration, it's necessary to perform dimensionality reduction to simplify the information [26], making it difficult to analyze the influence of various factors on production when the material flow is coupled with energy flow, thus offering limited guidance for optimization of the production.

Based on the Tecnomatix Plant Simulation development platform, this paper builds a coupled simulation model. The model considered the operation behaviors of the material flow and energy flow, the operation and technical model of the production equipment and the temperature-decreasing model of the ladle that reflects the energy flow. Different organizing strategies were employed, the performances were evaluated by the production indicators, i.e. ladle cycle time, the temperature of hot metal charged into BOF, operating cost.

2. Simulation model description

2.1 Model description

Driven by the time flow, the material flow and the energy flow in the manufacturing process converge in a certain production process where the processing is handled under the interaction of the two flows. The synergetic coupling of the two flows are reflected through the product output and the energy consumption. Therefore, the simulation model Mod, which consists of Plan P, Time τ, Equipment N and E, Operation f and Material Temperature Component M as shown in Eq. 1, realize the simulated expression to the real manufacturing processes by abstracting their physical characteristics and running rules, as shown in Eq. 2.

$$Mod = f(N, M, E, P^{\tau}) \tag{1}$$

$$S^{\tau} = \begin{cases} f_N(Grid, ME, TE) \\ f_M(Ent, Con) \\ f_E(T_{tra}, AE_{pro}, AE_{Equ}) \\ f_{P^{\tau}}(Equ, Cra, SC) \end{cases} \tag{2}$$

In Eq. 2, S^{τ} refers to the system status at time τ. $f_N(Grid, ME, TE)$ is the modeling rule whereby the ME (main equipment), TE (transportation equipment) and the directed $Grid$ constitute a model operation network. $f_M(Ent, Con)$ describes the operation rules on how the positions and properties change of the material entities Ent (hot metal) and the containers Con (ladle). $f_E(T_{tra}, AE_{pro}, AE_{Equ})$ expresses the energy flow rules. According to the roles of the energy played in the production, the rules contain the energy changing rule T_{tra} characterized in temperatures during transportation between processes, auxiliary energy adding rules AE_{pro} characterized by the addition of auxiliary materials during the processing, and energy consuming rules AE_{Equ} characterized by the consumption of materials when equipment is running. $f_{P^{\tau}}(Equ, Cra, SC)$ is the production (operation) rules at time τ of the equipment or transportation devices (crane Cra and span car SC) under a production schedule.

Through its four simulation modules on logistics, equipment, control and information, as shown in Fig. 2, the simulation model above can simulate the interaction and conversion process of the material and energy during the ladle circulation in a way much closer to the reality:

1) When modelling, the equipment module is first selected to build the simulated production environment. The equipment parameters set include the operating procedures, durations, types and amounts of energy consumed;
2) When running, the logistics rules in the operation control module and the control models in the material-energy control module can be called to push the simulation running as planned and calculate the changes in transportation temperature, material and energy medium consumption occurred during the ladle circulation. Various process data is recorded by the information module, which realizes the quantitative analysis of the simulation system.

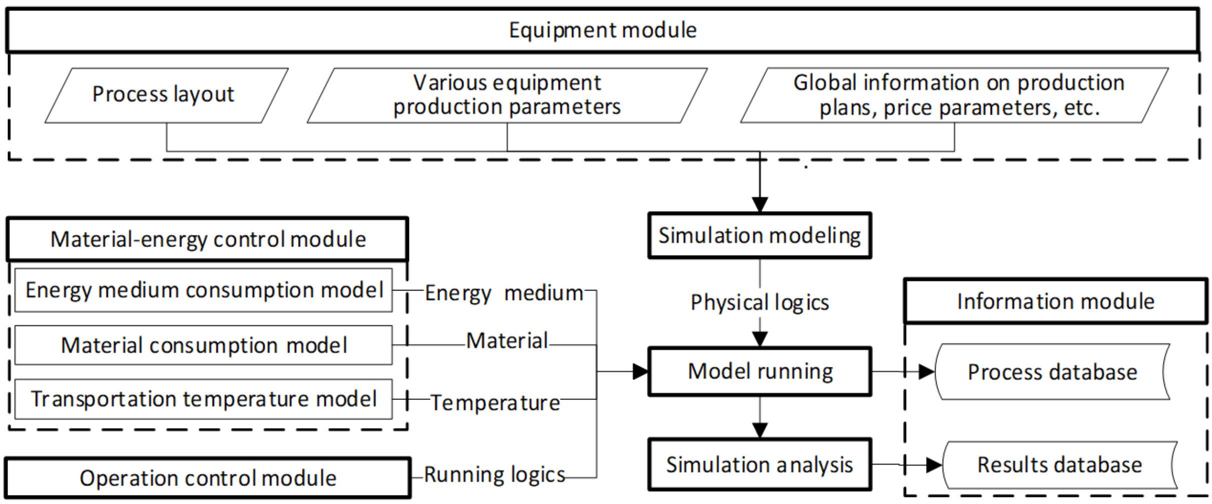


Fig. 2 Modules in simulation model

2.2 Functions of two key modules

The control modules in the simulation model contain the operation control module and the material-energy control module, as shown in Table 1, which serve as the running brain of the model, directing the orderly changing of the materials and energy in the logistical network. The operation control module controls the orderly transportation and processing of the material units between or in the process(es); while the material-energy control module is responsible for calculating the energy media needed to achieve the target temperature.

Table 1 Technical parameter design for two key modules

Type	Model		Description	Function
Operation control module	Production plan model		P_{TPush}	Decide the production plans, target composition and temperatures for each process according to the casting plan
	Crane operation and scheduling model		C_{Sche}	Order the cranes to complete the transportation tasks
	Span car operation and scheduling model		V_{Sche}	Order the span car com transporting task across bays
	Process equipment operation model		E_{Sche}	Executing production task according to equipment operating procedures
Material- energy control module	Material consumption model	Production specification model	M_{Rule}	Set the consumption per ton of steel in a certain process
		Linear consumption model	M_{Line}	Material consumption has a linear relationship with the components to be processed
		Non-linear relational model	M_{NonL}	Material consumption has a linear non-relationship with the components to be processed
		Temperature compensation model	M_{TCom}	Adopt physical heating method to reach target temperature
	Energy medium consumption model	Electric energy consumption model	Q_{Power}	Electric energy consumption for equipment operation
		Water consumption model	Q_{Water}	Consumption of water of various types for equipment operation
		Gas consumption model	Q_{Air}	Consumption of gases for equipment operation
	Temperature change model	Temperature drop model for hot metal charged into different ladles	T_{Charg}	Temperature drop caused by change of the hot metal containers
		Temperature drop model for hot metal transportation	T_{Trans}	Temperature changes of hot metal under different working conditions
		Temperature change model for the lining of an empty ladle	T_{Ladle}	Temperature changes of empty ladle under different working conditions

2.3 Logic flow of simulation operation

The actual production can be deemed as the combination of a series of chronological events, including the transport events and processing events of the unit material. Therefore, the coupling simulation model employs the trigger-based discrete simulation mechanism in which the temperature and material composition information carried by the hot metal is set as the conditions to activate various sub-models of the control modules under different events. The activating conditions are shown below:

Start Simulation
  Call P_{TPush} for production plan
Do
    Scan task queue
    Select task characteristics
    Case transportation task
        If Crane transportation then
      Call C_{Sche} and T_{Charg} and T_{Trans} and T_{Ladle}
      End
      If Span car transportation then
        Call V_{Sche} and T_{Charg} and T_{Trans} and T_{Ladle}
      End
    Case process processing task
      Call E_{Sche}
      Call M_{Rule} and M_{Line} and M_{NonL} and M_{TCom}
      If using power, then
          Call Q_{Power}

```
            End
            If using water, then
                Call Q_Water
            End
            if using gaseous media, then
                Call Q_Air
            End
        End inspect
    Loop until Simulation End
```

In the simulation operation, the temperature of hot metal is used as the condition to decide the priority of related transportation tasks. When reaching the process equipment, the temperature and composition of hot metal is firstly corrected according to the transportation situation and then treated as the inputting condition for the processing tasks. The corresponding material-energy model is enabled to calculate the materials and energy added to bring the temperature and property of the hot metal to reach the target. After that, the new information on hot metal temperature and composition is achieved and a loop starts until the end of the final process. More details are shown in Fig. 3.

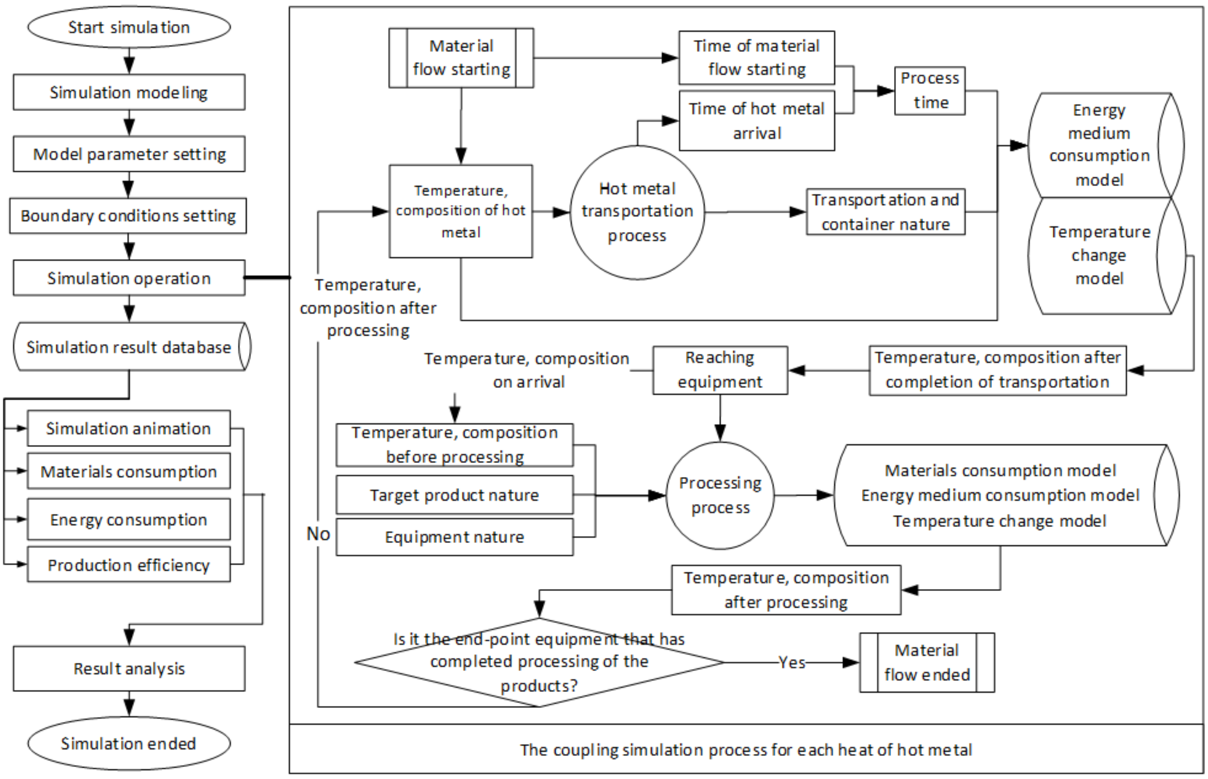


Fig. 3 Model operation flow

3. Results and discussion: Case studies

3.1 Simulation cases

The advantages the One-Ladle Technique offers vary greatly with the enterprises where it was employed. Aiming at offering optimal recommendations to improve the ironmaking-steelmaking interface logistics of Steelworks, A using One-Ladle Technique, a simulation model is built to obtain the operation solutions under different production strategies. Fig. 4 shows the model built on the operating relationships between specific equipment, material flow and energy flow of Steelworks A. The interface is composed of 2 BF, 4 KR, 3 BOF and other auxiliary devices. The ladle is transported by the span car between BF and buffer zone, and by crane between buffer zone, KR and BOF. The moving path of the full ladle is marked by the directed red lines while the path of the empty ladle by the directed green lines.

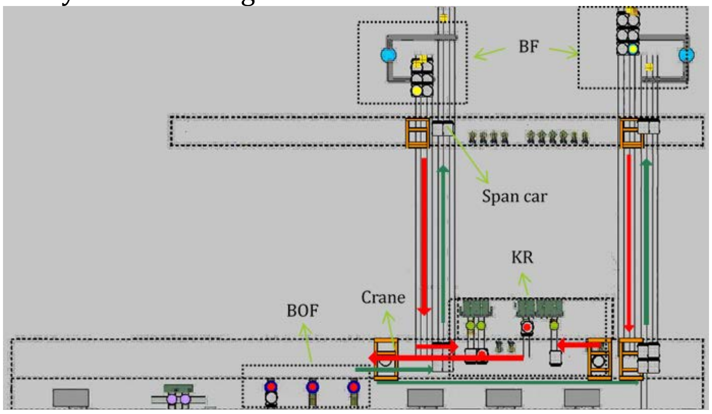


Fig. 4 Model simulation operation

In Fig. 4, each circle represents an empty ladle, and each red filled circle, and yellow filled circle stands for a ladle fully charged and a ladle half charged. When a filled circle is in the center of the equipment, it means a ladle is under processing at that station. By setting different process parameters in the model, the simulation operation provides indicative parameters describing the operating characteristics of the system quantitatively, allowing for analyses on how to ensure optimization of the production with the One-Ladle Technique.

According to the historical performance record of the two BFs in Steelworks A, 4 to 5 ladles on average are prepared to charge hot metal from the two under normal conditions, and 6 ladles are used at the peak period. Each BF has two tapping outputs through which the hot metal is tapped, the ladles prepared and tapping lines of each BF are set as shown below in Table 2.

In real operation, the tapping durations of the two BFs is different. Therefore, two tapping conditions are considered for #1 BF and #2 BF: One is that both BFs tap at the same time, and the other is that one taps 1 hour earlier than the other. These two tapping conditions, when combined with the ladle preparation patterns as mentioned above in Table 2, lead to 6 different working conditions, as shown below in Table 3.

In order to analyze the impact of different logistics parameters on production, the simulation is modeled according to the 1:1 layout of the ironmaking-steelmaking interface of Steelworks A. The simulation span is 24 h, during which 92 heats of hot metal are transported into the BOF. The hot metal output rate of the BFs is set to be 400 t each time, the hot metal tapping rate is 5 t/min, the initial temperature of the hot metal is 1500 °C. The BOF processing time is 38 min/ while the KR processing time 36 min. The speed of the span car carrying a full ladle is 20 m/min, and 30 m/min when carrying an empty ladle. The running speed of the crane is 60 m/min and the hoisting/dropping time 2.5 min. The temperature drop in the transportation process is 1 °C/min.

Table 2 Number of ladles prepared for each tapping line

		Ladle prep. pattern 1	Ladle prep. pattern 2	Ladle prep. pattern 3
Number of tappings per day		12	12	13
Total number of ladles made available per BF		6	5	4
Taphole A	Number of tapping lines	2	2	2
	Number of pallets per line	3	3	3
	Number of ladles for #1 tapping line	3	3	2
	Number of ladles for #2 tapping line	3	2	2
Taphole B	Number of tapping lines	2	2	2
	Number of pallets per line	3	3	3
	Number of ladles for #3 tapping line	0	0	0
	Number of ladles for #4 tapping line	0	0	0

Table 3 Establishment of simulation cases

Case No.	Ladle preparation pattern	Tapping condition
Case I	Pattern 1	Two BFs tapping at the same time
Case II	Pattern 1	One BF tapping 1 hour earlier than the other
Case III	Pattern 2	Two BFs tapping at the same time
Case IV	Pattern 2	One BF tapping 1 hour earlier than the other
Case V	Pattern 3	Two BFs tapping at the same time
Case VI	Pattern 3	One BF tapping 1 hour earlier than the other

3.2 Analyses of the simulation model results

Analyses of the simulation model results are included in the comprehensive evaluations of the material flow and the energy flow. Therefore, the ladle cycle period ρ_{ladle} is used to evaluate the material circulation efficiency (see Eq. 3). It shows that when the proportion of heats is higher while the cycle time for those heats is shorter, the material flow operation efficiency is higher. And, the average temperature of the hot metal charged into BOF t_{charge} is adopted to analyze the energy losses of the system (see Eq. 4). It suggests that the higher the average temperature of hot metal charged into BOF, the lower the temperature drop of the system. The operating costs C comprehensively reflects the situation of various material energy consumed by the system under the action of the material flow coupled with the energy flow, which is useful to evaluate the production outcomes (see Eq. 5).

$$\rho_{ladle} = \frac{Num(T_{min} \leq T_{ladle} < T_{max})}{n} \times 100\ \% \tag{3}$$

$$t_{charge} = \sum_{i=1}^{n} t_i\ /n \tag{4}$$

$$C = \sum_{j=1}^{m} (P_j \times W_j)\ /S_{weight} \tag{5}$$

In these equations, $Num(T_{min} \leq T_{ladle} < T_{max})$ represents the number of heats within the ladle cycle time, between T_{min} and T_{max}; n means the total number of heats; t_i is the temperature of the hot metal in a heat when it can be charged into the BOF; P_j refers to the unit price of the Material j consumed; W_j stands for the weight of the Material j consumed; m is the total types of materials consumed, and S_{weight} is the amount of the molten steel produced.

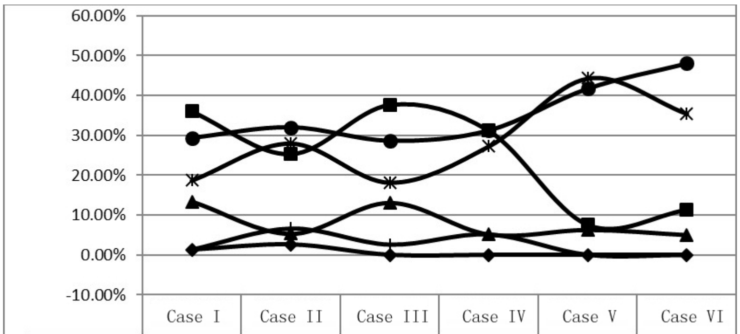
3.3 Analyses of simulation results

The simulation test was carried out according to the aforementioned input conditions. From the perspective of material flow operation efficiency, production condition is a significant impact factor. As shown in Fig. 5, there is a significant difference in the distribution of the ρ_{ladle} of each case. Both Cases I and Ⅱ have a production situation where the ladle cycle period is greater than 6 hours. This means that some ladles are in a condition waiting for the hot metal tapping; the logistics operation efficiency of Case VI is the highest, and the ratio of the number of furnaces in the range of 0 to 2 hours for the turnover time of the iron ladle is 48.11 %, which is the lowest (case 1) 18.78 % higher. It can be seen that the tank allocation system 3 and the interval tapping of the two BFs are beneficial to improve the operating efficiency of the material flow.

Temperatures of the hot metal charged into the BOF are used as an indicator to measure the energy flow conversion. Under the same initial temperature conditions, the higher the temperature of the hot metal transported into BOF, the smaller the temperature drop of the system. Fig. 6 presents the effect of different material flow operation efficiencies on the energy flow. The average temperature of the hot metal charged into the BOF in Case VI with the highest material flow operation efficiency is 10.3 °C higher than that in Case I. This indicates that the higher the material flow operating efficiency, the smaller the temperature drop of the system.

Different transportation time of the hot metal causes temperature fluctuation when it arrives at the BOF, thus influencing the consumption of production materials. In the simulation, the material consumption model is used to calculate the consumption of material and energy for each process realization of equal changes in the nature of hot metal material flows at different temperatures, and the production effects are comprehensively reflected in the operating costs. Fig. 7 shows the comparison of operating costs between different cases by using Case I as the benchmark. The costs per tonne of steel in Case VI are CNY 1.02 yuan lower than in Case I, which is an annual CNY 10.2 million yuan saving considering a steel plant with a 10 million tonnes capacity of the steel production.

The built simulation model is, as shown above, able to reproduce the processes in which the material flow and the energy flow interact at the ironmaking-steelmaking interface. It also can quantitatively reflect the impact of different production strategies on operation in terms of indicators such as ladle cycle time, temperature of hot metal charged into BOF, and operating costs, thus eliminating the restrictions of evaluating production only from the perspective of material flow efficiency or temperature control in energy flow. The analyses of the simulation results also indicate that application of the optimization strategies, e.g., increasing the tapping frequency, reduces the total number of ladles prepared for each BF. Moreover, an 1 hour tapping interval of two BFs enables Steelworks A to increase the average temperature of the hot metal charged into BOF by 10.28 °C, and reduce the cost per tonne of steel by CNY 1.02.



	Case I	Case II	Case III	Case IV	Case V	Case VI
T>6h	1.33%	2.67%	0.00%	0.00%	0.00%	0.00%
5h<T≤6h	1.33%	6.67%	2.60%	5.19%	0.00%	0.00%
4h<T≤5h	13.33%	5.33%	12.99%	5.19%	6.33%	5.06%
3h<T≤4h	36.00%	25.33%	37.66%	31.17%	7.59%	11.39%
2h<T≤3h	18.68%	28.00%	18.18%	27.27%	44.31%	35.44%
0h<T≤2h	29.33%	32.00%	28.57%	31.18%	41.77%	48.11%

Fig. 5 Distribution of the ρ_{ladle} of each case

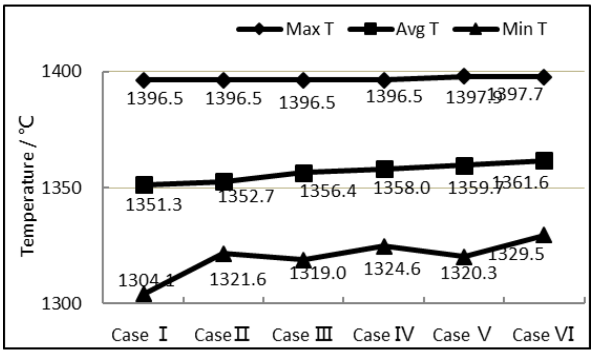


Fig. 6 Comparison of temperatures of hot metal charged into BOF

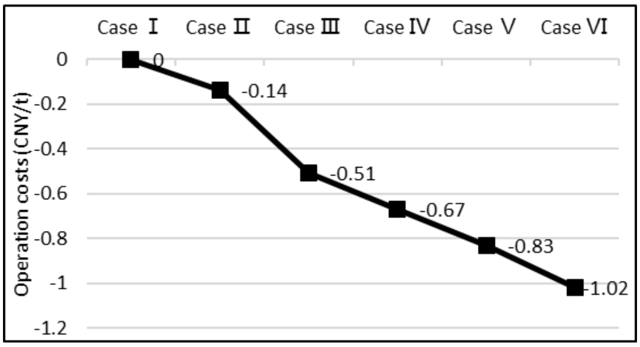


Fig. 7 Comparison of operating costs (in CNY per tonnes of steel)
Note: The operating cost in Case I is used as the benchmark reference

4. Conclusion

The coupling simulation model built in the study considered the operation behaviors of the material flow and energy flow of the ironmaking-steelmaking interface, the operation and technical model of the production equipment and the temperature drop model of the ladle. The energy-flow-driven moving process of the material simulated by the model reflects the interrelated but independent relation between material flow and energy flow.

The simulation results suggest that the simulation model can provide comprehensive decision-making supports for actual operation control and optimization of production management strategies. The optimized production management strategies result in an increase of 10 °C in the average temperature of the hot metal charged into BOF and a reduction in the cost per tonne of steel by CNY 1.02.

The methodology for building the coupling simulation model can be extended to the entire steel manufacturing process and become an optimizer to steel producers by providing decision-making supports in achieving efficiency improvements and cost reductions.

Acknowledgment

This work was financially supported by the National Natural Science Foundation of China (Grant No. 51734004) and partly by the National Key Research and Development Program of China (Grant No. 2020YFB1712803 and No. 2017 YFB0304005).

References

[1] Zhao, J.Y., Wang, Y.J., Xi, X., Wu, G.D. (2017). Simulation of steel production logistics system based on multi-agents, *International Journal of Simulation Modeling*, Vol. 16, No. 1, 167-175, doi: 10.2507/IJSIMM16(1)CO4.

[2] Wang, X., Wei, Z., Wang, S., Zhou, W., Liu, S. (2012). Multi-agent based simulation system modelling for steelmaking-continuous casting production process, In: *Proceedings of 2012 24th Chinese Control and Decision Conference (CCDC),* Taiyuan, China, 170-173, doi: 10.1109/CCDC.2012.6242928.

[3] Fan, X.-K. (2010). Hot metal transportation method and general layout of iron & steel plant, *Architectural and Structural Design*, No. 10, 29-34.

[4] Li, M.-Y., Yan, Y.-G., Lu, S.-W., Luo, Z.-G. (2016). Process design and simulation of iron-steelmaking zero interface, *Shandong Metallurgy*, Vol. 38, No. 4, 26-29.

[5] Shang, G., Xiang C., Duan, X. (2017). Research and application of general layout parttern of "one-ladle mode" on ironmaking/steelmaking interface, *China Metallurgy*, Vol. 27, No. 8, 45-49.

[6] Wiyaratn, W., Watanapa, A. (2010). Improvement plant layout using systematic layout planning (SLP) for increased productivity, *International Journal of Mechanical, Industrial and Aerospace Sciences,* Vol. 4, No. 12, 1382-1386, doi: 10.5281/zenodo.1330985.

[7] Penteado, F.D., Ciric, A.R. (1996). An MINLP approach for safe process plant layout, *Industrial & Engineering Chemistry Research*, Vol. 35, No. 4, 1354-1361, doi: 10.1021/ie9502547.

[8] Balakrishnan, J., Cheng, C.-H., Wong, K.-F. (2003). FACOPT: A user friendly facility layout optimization system, *Computers & Operations Research*, Vol. 30, No. 11, 1625-1641, doi: 10.1016/S0305-0548(02)00096-5.

[9] Qiu, J., Tian, N.-Y. (2005). Relative superiority research on the ironmaking/steelmaking interface of the typical process section, Journal of University of Science and Technology Beijing, Vol. 27, No. 6, 740-744.

[10] Zhang, L.-Q., Tian, N.-Y., Xu, A.-J. (2008). Selection of hot metal pretreatment mode for new steel-making plant,

Steelmaking, Vol. 24, No. 1, 58-62.

[11] Tang, L., Wang, G., Liu, J. (2007). A branch-and-price algorithm to solve the molten iron allocation problem in iron and steel industry, *Computers & Operations Research*, Vol. 34, No. 10, 3001-3015, doi: 10.1016/j.cor.2005.11.010.

[12] Wang, G., Tang, L. (2007). A column generation for locomotive scheduling problem in molten iron transportation, In: *Proceedings of 2007 IEEE International Conference on Automation and Logistics,* Jinan, China, 2227-2233. doi: 10.1109/ICAL.2007.4338946.

[13] Yang, F., Shi, M. (2019). Emergency surgery scheduling under urban emergencies based on improved moth-flame optimization, *Journal Européen des Systèmes Automatisés,* Vol. 52, No. 1, 49-55, doi: 10.18280/jesa.520107.

[14] Su, J., Wei, M., Liu, A. (2018). A robust predictive–reactive allocating approach, considering random design change in complex product design processes, *International Journal of Computational Intelligence Systems,* Vol. 11, No. 1, 1210-1228, doi: 10.2991/ijcis.11.1.91.

[15] Bedhief, A.O. (2021). Comparing mixed-integer programming and constraint programming models for the hybrid flow shop scheduling problem with dedicated machines, *Journal Européen des Systèmes Automatisés,* Vol. 54, No. 4, 591-597, doi: 10.18280/jesa.540408.

[16] Su, J., Yang, Y., Yu, K., Zhang, N. (2018). A method of partner selection for knowledge collaboration teams using weighted social network analysis, *Journal of Intelligent Systems,* Vol. 27, No. 4, 577-591, doi: 10.1515/jisys-2016-0140.

[17] Xiao, P., He L. (2014). Applications of logistical simulation to one-ladle-mode iron-making/steel-making interface, *Steelmaking*. Vol. 30, No. 4, 70-74.

[18] Xia, W., Song, W., Xia, Q. (2015). The application of system simulation and logistics automation technology in the steel production logistics process, In: *Proceedings of the 2015 International Conference on Advanced Material Engineering,* Guangzhou, China, 622-627, doi: 10.1142/9789814696029_0072.

[19] Han, W.-G., Li, X.-P., Shi, Y.-X, Wang, W.-D., Zhang, C.-X. (2013). Online ladle quantity of "one-open-ladle-from-BF-to-BOF" route by using queuing theory, *Iron & Steel,* Vol. 48, No. 5, 21-24.

[20] Huang, B.-F., Shi, Z., Zhou, X.-L., ZHANG, G.-F. (2015). Control model of torpedo ladle based on hot metal supply and demand, *Journal of Kunming University of Science and Technology（Science and Technology),* Vol. 40, No. 6, 133-138.

[21] Du, T., Cai, J.J., Li, Y.J., Wang, J.J. (2008). Analysis of hot metal temperature drop and energy-saving mode on techno-interface of BF-BOF route, *Iron & Steel,* Vol. 43, No.12, 83-86.

[22] Ren, Y.-J., Wang, J.-W., Zhang, X.-B., Zhao, H.-W. (2012). Prediction model of hot metal temperature for BF-BOF interface based on LM BP neural network, *Iron & Steel*, Vol. 47, No. 9, 40-39.

[23] Liu, S.W., Yu, J.K., Yan, Z.G., Liu, T. (2010). Factors and control methods of the heat loss of torpedo-ladle, *Journal of Materials and Metallurgy,* Vol. 9, No. 3, 159-163.

[24] Chen, D.M., Liu, Y.H., He, S.F., Xu, S., Dai, F.Q., Lu, B. (2020). Fuel gas operation management practices for reheating furnace in iron and steel industry, *Advances in Production Engineering & Management,* Vol. 15, No. 2, 179-191, doi: 10.14743/apem2020.2.357.

[25] Chen, D., Lu, B., Chen, G., Yu, W. (2017). Influence of the production fluctuation on the process energy intensity in iron and steel industry, *Advances in Production Engineering & Management,* Vol. 12, No. 1, 75-87, doi: 10.14743/apem2017.1.241.

[26] Xiao, Y., Li, C., Song, L., Yang, J., Su, J. (2021) A multidimensional information fusion-based matching decision method for manufacturing service resource, *IEEE Access,* Vol. 9, 39839-39851, doi: 10.1109/ACCESS.2021.3063277.

APEM journal

Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 359–371
https://doi.org/10.14743/apem2021.3.406

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

Recharging and transportation scheduling for electric vehicle battery under the swapping mode

Huang, A.Q.[a,*], **Zhang, Y.Q.**[b], **He, Z.F.**[a,*], **Hua, G.W.**[b], **Shi, X.L.**[b]

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

A B S T R A C T

Electric vehicle battery recharging on the swapping mode has grown up as an important option other than the plug-in recharging mode in China, given that several auto giants have been dedicated in constructing their battery swapping systems. However, the lack of effective operational methods on battery recharging and transportation scheduling has aroused a big challenge on the practical application of the swapping mode, which enables the necessity of our work. This study proposes a joint optimization model of recharging and scheduling of electric vehicle batteries with a dynamic electricity price system which is able to identify the optimal charging arrangement (the recharging time and the quantity of recharging batteries) as well as the optimal transportation arrangement (the transportation time and the quantity of transporting batteries). For the validation purpose, a numerical study is implemented based on dynamic electricity prices in Beijing. A sensitivity analysis of parameters is carried out to increase the robustness and provide more managerial insights of the model.

A R T I C L E I N F O

Keywords:
Electric vehicle;
Battery recharging;
Battery swapping;
Battery logistics;
Transportation scheduling

**Corresponding author:*
aqhuang@bjtu.edu.cn
(Huang, A.Q.)
hezefang@bwu.edu.cn
(He, Z.Q.)

Article history:
Received 17 August 2021
Revised 24 October 2021
Accepted 28 October 2021

1. Introduction

Carbon emission reduction is becoming a privilege for an increasing number of countries across the world to counter global warming effects. At the Climate Ambition Summit in December 2020, the leaders of the 27 EU member states stated that by 2030 their net greenhouse gas emissions would be reduced to no more than 45 % of 1990. China has also made a commitment to peak carbon dioxide emissions before 2030 and achieve carbon neutrality before 2060. Given that automobile exhaust is one of the major sources of carbon emissions while electric vehicles (EVs) are clean, the use of EVs has considerable potential in reducing vehicle emissions [1-5]. Therefore, the extensive application of EVs is an important available measure to help the countries fulfill their commitments. Furthermore, for a country, some pillar industries like manufacturing could enjoy extra emission credits saved by the transportation sector using EVs, and thus have more development vigor.

However, the range anxiety of consumers has severely limited the popularity of EVs. It usually takes hours for EVs to recharge an EV, which is a matter of great concern for consumers, in addition to the imperfect recharging infrastructure [6, 7]. According to the Development Strategy Report of the Ministry of Industry and Information Technology, the number of EVs in China will reach 60 million by 2030. If a large scale of EVs is charged in a disordered way (randomly connected to the power grid for charging), it will bring damages to the power grid [8]. Different from in the plug-in recharging mode, the depleted battery of an EV in the swapping mode is replaced with a full one, then the EV could continue to run and the depleted battery is left to be recharged. The swapping mode separates the recharging process from the battery swapping, which enables three advantages: (1) the whole operation takes less than 10 minutes, which is much faster than plug-in recharging; (2) the charging time and the quantity of charging batteries can be scheduled reasonably so as to reduce the impacts on the safety and quality of power grid [9-10]; (3) the depleted batteries can be charged during off-peak hours of a discounted electricity price and consequently reduce the charging cost. Actually, the swapping mode has been booming in China, e.g., NIO, a Chinese giant EV-maker, has set up 139 swapping stations, BAIC Group made a plan to construct 100 swapping stations capable to serve no less than 10,000 EVs, Changan Auto set up a battery swapping alliance to promote the swapping business, etc.

Although the key challenges for operating battery swapping mode is to optimize both the charging time of depleted batteries and the battery transportation scheduling between battery swapping station (BSS) and battery charging station (BCS) [11], current studies on the battery swapping mode rarely address the joint optimization of the above two problems, not to mention the joint problem with the dynamics of electricity price. To bridge this gap, we study a joint optimization problem of charging time and transportation scheduling based on a battery swapping and charging system (BSCS). Although the BSCS hereby has a simple structure to comprise one BSS and one BCS, it is able to unveil the nature of the swapping and charging systems and could be easily extended to complex ones. This study aims to answer the following two questions: (1) how to determine the optimal charging time and quantity of the charging batteries; (2) how to optimize the transportation scheduling of batteries between BCS and BSS. All abbreviations are detailed in Table 7 in Appendix.

The uniqueness of this work is multifold. First, the proposed model takes into account the dynamic electricity prices, which not only maintains the security of the grid but also reduces charging costs. Secondly, this work is the first to address the joint optimization problem of the centralized charging and the transportation scheduling of batteries with consideration of the dynamic electricity price, which could identify the optimal charging time of the depleted batteries, as well as the optimal time and quantity of the battery transportation between BSS and BCS. Finally, we highlight the managerial implication by carrying out a sensitivity analysis.

The remainder of this paper is organized as follows. Section 2 reviews the latest studies on the battery swapping mode. Section 3 elaborates the problem and the mathematical model. Section 4 shows the details of a numerical study and a sensitivity analysis, based on which, Section 5 provides managerial insights. Section 6 concludes this study and discusses some potential future directions.

2. Literature review

Substantial studies have been made on the charging strategies of EVs [12-15], however, there are also many insurmountable challenges on the charging strategies, such as long charging time, charging inconvenience, etc. In recent years, with the invention of battery swapping station, the battery-swapping mode has been obtaining increasingly more attention from industries and the academia [16].

Most of the studies on the battery swapping strategy focus on the optimal charging time of batteries. Kang *et al.* [17] proposed a novel centralized charging strategy of EVs under the battery swapping scenario based on spot electric price and design a population-based heuristic approach to minimize total charging cost while reducing power loss and voltage deviation of power networks. Zheng *et al.* [18] focused on EV battery swapping station coordinated charging dispatch method based on CS algorithm to achieve the optimization of daily charging plan of battery swapping station (BSS). Yang *et al.* [19] propose a dynamic operation model of BSS in electricity market based on the short-term battery management, and acquires additional revenue by responding actively to the price fluctuation in electricity market. Song *et al.* [20] designed a typical connection mode of electric vehicle charging and battery exchange infrastructure, which can provide guidance for the planning of electric vehicle charging and battery exchange infrastructure interconnected with the grid. Zhang *et al.* [21] proposed an optimized charging mode (OCM) to determine the impact of drivers' switching behavior on power grid and power generation cost. Wang *et al.* [22] proposed a comprehensive optimal allocation method for EV switching stations based on orderly charging strategy. Infante *et al.* [23] considered the demand of EV users and the load of power grid, and proposed a strategy model of EV switching station to optimize the benefits. Sarker *et al.* [24] proposed the optimization framework of the battery switching station operation model. Wang *et al.* [25] proposed an integrated optimization model with EV charging station, battery-swap station and energy storage system, which aims to find a balance status between the power grid and the EV users during the power flow exchange in the background of internet energy.

To the best of our knowledge, most of the existing studies focus on the charging process of the battery exchange system, while ignoring the transportation scheduling between BCSs and BSSs. Given that the operation of BSCS needs to coordinate the two perspectives, it is necessary to propose a joint optimization model to identify the optimal charging time and the optimal quantity of the charging battery, as well as the optimal battery transportation schedule (transportation time and transportation quantity) between BCS and BSS.

3. Model definition

3.1 Problem formulation

We study the battery charging process of the battery charging station (BCS) and the swapping process of the battery swapping station (BSS) in the battery swapping and charging system (BSCS). There are three subsystems in BSCSs: the battery swapping station (BSSs), the battery charging station (BCSs) and the logistics system between them. In this system, users send the depleted batteries (DBs) back to BSSs and remove the fully charged batteries (FBs). After a while, BSSs will ship the accumulated depleted batteries back to BCSs. The depleted batteries (DBs) will be charged centrally by BCSs within a reasonable time and then shipped back to BSSs for further use by future customers. Based on the above process, we propose a mathematical model of battery centralized charging and transportation scheduling. Furthermore, we assume that the service time of logistics system will not be discussed in our model.

We introduce a mathematical formulation of this problem. First, the model discretizes the continuous time according to a certain time granularity. According to the time-of-use electricity price mechanism and the uncertainty of demand, managers need to make decisions at the endpoints of each time grain. For BSS, the managers need to make decisions about when to transport how many batteries from BCS to meet the customer's demand for electricity conversion. For BCS, the decision is when to charge the depleted batteries, so as to meet the demand of BSS while minimizing the charging cost. Certain transportation costs will be incurred in the process of transportation. For the whole system, in order to minimize the total cost, on the basis of meeting the demand, when to transport and how many batteries to transport are also the contents to be decided.

The model makes the following assumptions:

1. On the BSCS side, we assume that there is only one BCS, which is responsible for charging the battery. There is only one BSS, which is responsible for battery replacement for users. The logistics process between BCS and BSS does not consider the transport time and route temporarily, we only study the number and quantity of transport.
2. In terms of electricity price, this section assumes that electricity price only change at the time granularity endpoint. According to the State Grid, electricity price is divided into three levels, peak, normal and trough, and only change at the hourly point of each day. The

electricity price fluctuates periodically in a stepped manner. Therefore, this setting is also realistic.

3. In terms of battery charging process, we assume that each battery will charge for the same time. In fact, the residual power of the battery is random and the time required for full charge should also be random. But EV users who generally go to BSS for energy renewal have a relatively low battery surplus. We generally set the initial SOC of depleted batteries as 0.2, and the SOC at full charge is between 0.9-1. Consider that the batteries currently used by car manufacturers are mostly made by BYD (mainly for BYD's own use) and Ning-de Times. So even if the car brand is different, the battery specifications are similar, which makes the time required for different batteries to be fully charged is basically the same, so this assumption will not lead to a greater deviation from reality.

3.2 Model specification

Without loss of generality, we characterize the scenario of our model as follows:

1. The battery belongs to BSS in BSCS;
2. There is only one BCS;
3. BSS cannot charge DBs;
4. The battery type is the same. DBs all have the same SOC, so does WBs, so they are charged at the starting point of time granularity, and all batteries have the same charging time;
5. The battery delivery time between BCS and BSS is not considered;
6. The total operating period is divided into discrete time periods. Let g represent the time granularity;
7. Electricity prices change over time and only at the end points of time granularity.
8. To extend battery life, the battery does not stop charging until it is fully charged.

It is notable that although the BSCS hereby comprise one BSS and one BCS, it is sufficient to reflect the nature of the swapping and charging systems and could be easily extended to complex ones. The natation of variables and parameters are shown in Table 1.

Table 1 Variables and parameters of the model

Variable	Meaning
t	t time granularity;
TP	Single vehicle single transport cost
P_t	Electricity price at the t time granularity
W_t	The number of full batteries in the BSS at the beginning of the t period
D_t	Power exchange demand at the beginning of t
Q_t^*	BCS full battery number at the beginning of t
Q_t	Number of batteries charged at the beginning of t
Q_t^+	Number of batteries charged in t period
TQ_t^{out}	Number of batteries shipped from BSS to BCS at the beginning of t
TQ_t^{in}	Number of batteries shipped from BCS to BSS at the beginning of t
T_t^{out}	Number of vehicles transporting batteries from BSS to BCS at the beginning of t
T_t^{in}	Number of vehicles transporting batteries from BCS to BSS at the beginning of t
Q_c	The maximum transport capacity of a single vehicle
N_{occup}	The ratio of full batteries to the capacity of the swapping station at initial time
N_c	The proportion of charging piles that are not occupied in the initial BCS
M	Number of time particles required for a single battery to be fully charged
C_{aps}	The maximum number of cells a BCS can hold
C_{apc}	The maximum number of cells a BCS can hold
PoC	Charging power
T	The total time window length of this study

The model:

$$Min\ S = \sum_{t=1}^{T}\left(g \cdot PoC \cdot Q_t \cdot \sum_{i=t}^{t+M-1} P_i\right) + TP\sum_{t=1}^{T}\left(T_i^{in} + T_t^{out}\right) \quad (1)$$

Subject to

$$W_1 = C_{aps} \times N_{occup} \tag{2}$$

$$W_t = W_{t-1} - D_{t-1} + TQ_{t-1}^{in} \tag{3}$$

$$W_t \geq D_t \tag{4}$$

$$Q_1^* = (1 - N_c) \times C_{apc} - TQ_1^{in} \tag{5}$$

$$Q_t^* = Q_{t-1}^* + Q_{t-1}^+ - TQ_t^{in} \tag{6}$$

$$TQ_t^{in} \leq Q_t^* \tag{7}$$

$$Q_t^+ \begin{cases} Q_{t-M+1}, t \geq M \\ 0, t \leq M \end{cases} \tag{8}$$

$$\sum_{i=1}^{t} Q_i \leq \sum_{i=1}^{t} TQ_i^{out} \tag{9}$$

$$\sum_{i=1}^{t} TQ_i^{out} \leq \sum_{i=1}^{t} D_i \tag{10}$$

$$\sum_{i=1}^{t} TQ_i^{in} - \sum_{i=1}^{t} TQ_i^{out} \leq C_{aps}\left(1 - N_{occup}\right) \tag{11}$$

$$\sum_{i=1}^{t} TQ_i^{out} - \sum_{i=1}^{t} TQ_i^{in} \leq N_c \cdot C_{apc} \tag{12}$$

$$\frac{TQ_t^{out}}{Q_c} \leq T_t^{out} \leq \frac{TQ_t^{out}}{Q_c} + 1 \tag{13}$$

$$\frac{TQ_t^{in}}{Q_c} \leq T_t^{in} \leq \frac{TQ_t^{in}}{Q_c} + 1 \tag{14}$$

$$t, P_t, W_t, Q_t^*, Q_t, Q_t^+, TQ_t^{in}, TQ_t^{out}, T_t^{in}, T_t^{out} \geq 0 \tag{15}$$

Objective function Eq. 1 is the minimum sum of battery charging cost and transportation cost. Specifically, the total cost of BSCS consists of two parts: (1) The electricity cost for battery charging under the time-of-use tariff mechanism; (2) The transportation cost of batteries dispatched between BCS and BSSs.

The constraints can be divided into 4 categories, including:

Constraints on the number of fully charged batteries in BSS

Constraint Eq. 2 indicates that the number of full batteries in the initial BSS is equal to the number of available batteries placed in the initial BSS.

Constraint Eq. 3 indicates that the number of full batteries in BSS in period t is equal to the sum of the number of full batteries in BSS in period $t-1$ and the number of batteries shipped into BSS, minus the number of batteries required in period $t-1$.

Constraint Eq. 4 indicates that the electrical changing demands that can be met in each phase should not exceed the total amount of fully charged batteries available in the current phase of the electrical changing station.

Constraint Eq. 11 is the inventory capacity constraint in BSS, indicating that the difference between the number of batteries shipped in and out of BSS in period t should not exceed the inventory capacity in BSS in period $t-1$.

Constraints on the number of fully charged batteries in BCS

Constraint Eq. 5 The number of fully charged cells in the initial BCS is equal to the number of available cells placed in the initial BCS minus the number of fully charged cells shipped out of the first BCS.

Constraint Eq. 6 indicates that the number of fully charged cells in BCS at period t is equal to the sum of the number of fully charged cells in BCS at the beginning of period $t-1$ and the number of fully charged cells at the beginning of period $t-1$ minus the number of cells shipped out of BCS at the beginning of period t.

Constraint Eq. 7 indicates that the number of batteries shipped out of BCS in each period should not exceed the total amount of fully charged batteries available in the charging station in the current period.

Constraint Eq. 12 is the constraint of BCS internal free charging capacity, indicating that the difference between the number of batteries transported in and out of BCS in period t should not exceed the spare charging capacity in BCS in period $t-1$.

Battery charge time constraint

Constraint Eq. 8 indicates that when the charging time of the battery is greater than or equal to the time particle size of M, the battery is fully charged.

Battery transport constraints in logistics system

Constraint Eq. 9 indicates that the number of batteries charged from the initial period to the current period should not exceed the total number of empty batteries in each BCS period.

Constraint Eq. 10 indicates that the number of batteries shipped out of BSS from the initial period to the current period should not exceed the demand for changing electricity.

Constraints Eq. 13 and Eq. 14 are vehicle constraints for transport.

Constraint Eq. 15 indicates that all variables are integers not less than 0.

This is a typical nonlinear mixed integer programming problem that can be solved by mathematical softwares such as MATLAB, CPLEX, and LINGO.

4. Numerical study and sensitivity analysis

To illustrate the effectiveness of the model, this section is based on the time-of-use electricity price mechanism of the power grid, and parameters are set according to the actual situation. We solve the problem by using LINGO mixed integer programming solver. In order to show the quantitative relationship among demand distribution, initial battery storage ratio and optimal charging time, we study the influence of demand distribution and initial battery storage ratio on the centralized charging strategy.

4.1 Parameter settings

- g = 0.5: The time particle size is 0.5, indicating that a decision should be made every 0.5 hour in the charging station, including whether to start charging a new batch of empty batteries and whether to transport the fully charged batteries to the changing station;
- M = 10: Suppose it takes 5 hours to fully charge a battery. That is 10 time granularities;
- C_{aps} = C_{apc} = 400: BSS and BCS can hold up to 400 batteries;
- N_{occup} = 0.8: Initially, 20% of the inventory capacity of BSS is idle, which means 80% of the existing inventory is occupied by full batteries;
- N_c = 1: There is no battery in the initial state of BCS;
- Q_c = 50: The maximum transport capacity of a single vehicle is 50 batteries;
- TP = 200: A transportation cost is 200 yuan;
- PoC = 3: The charging power is 3 kWh;
- $D\sim Passion(15)$: Referring to the average visit times of gas stations in real life, the demand is set to follow the Poisson distribution with an average of 15;

- P: According to the peak hours and corresponding electricity prices published by China Power Grid. The specific setting of electricity price is shown in Table 2:
- T: The research time is set as the time for all the initial fully charged batteries in BSS to be swapped out plus an integer multiple of M, combined with the above parameters, $a\ research\ cycle\ = \ 21 + 3 \cdot M$, considering the continuity of the model for the study cycle and repeatability, will postpone back 10 research cycle time granularity, namely the total time of this study window set to $21 + 3 \cdot M + 10$, the start time of study based on real life set to 7:00 in the morning.

Table 2 Electricity price distribution in different periods

Electricity Intensity	Period of time	Electricity price (yuan/kWh)
Peak	7:00-11:00	1.234
Normal	11:00-19:00	0.856
Peak	19:00-23:00	1.234
Trough	23:00-7:00（Next day）	0.376

4.2 Model solving

This part assumes that the electrical changing demand D follows the Poisson distribution with an average value of 15, and a group of samples with a capacity of 61 are randomly sampled. According to the electrical changing demand represented by this group of samples, the optimal charging time and quantity distribution as well as the battery transportation time and quantity distribution between the electrical changing station and the charging station are solved.

Fig. 1 shows the quantity distribution of electrical changing demands represented by this group of samples.

According to the determination principle of total time window length, the study cycle is 61 time granularity. The optimal solution under this sample electrical changing demand is shown in Fig. 2.

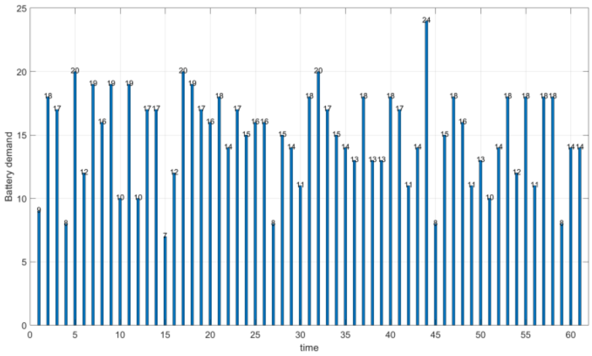


Fig. 1 Electrical changing requirements for each time granularity during the study period

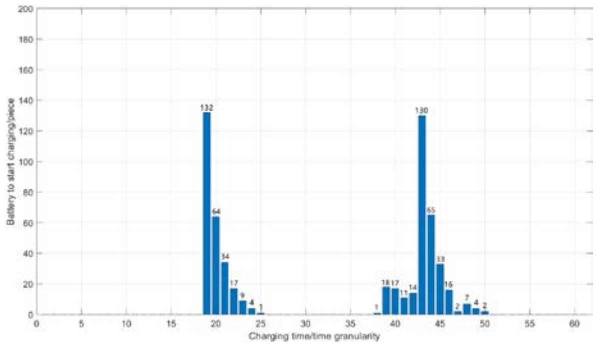


Fig. 2 Optimal charging time and quantity distribution of BCS

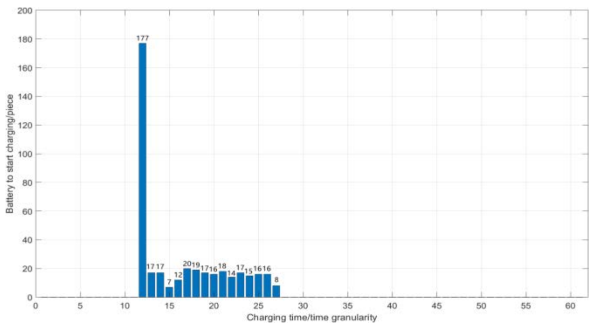


Fig. 3 Optimal time and quantity distribution of batteries from BSS to BCS

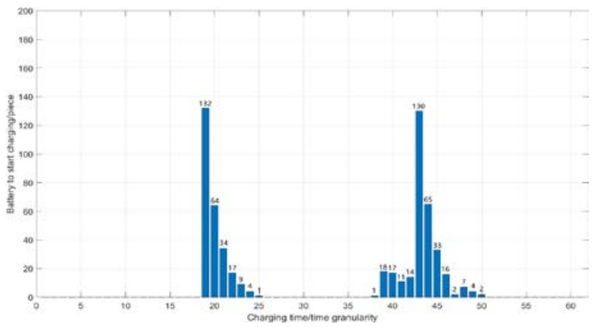


Fig. 4 Optimal time and quantity distribution of the battery from BCS to BSS

In the scenario of EV energy update set in this article, the following conclusions can be drawn:

- The optimal charging time and quantity distribution are shown in Table 3. Obviously, most batteries avoid the peak electricity consumption period and choose to start charging when the electricity price is normal or low.
- Table 4 shows the distribution of transportation time and quantity of batteries between BCS and BSS. We can found that the battery transportation time is relatively concentrated. As shown in Tables 4, the battery transportation presents asymmetry. The amount of transportation from BSS to BCS is greater than that in the opposite direction. The reason is that batteries are transported between BCS and BSS, and some batteries are still placed in BSS when the research time is cut off. Therefore, although there are inconsistencies in battery transportation in this experiment, it does not violate business logic and actual operation. The continuous operation of charging and changing station can be regarded as the periodic repetition of the time period studied in this paper. Therefore, we can use the same method to make optimization decisions regularly, so as to ensure the minimum charging and transportation costs under the premise of meeting the electrical changing needs in each cycle.
- As shown in Figs. 3 and 4, in the optimal time and quantity distribution of battery transportation between BSS and BCS, the number of batteries to be transported in the decision of the individual time granularity is very small. The research results show that the number of transport batteries is less than 10 in some cases, especially when the batteries are transported from BCS to BSS, the minimum number of transport is even one or two. In fact, the transportation cost is generally determined by the number of trips the vehicle takes, regardless of whether the vehicle is fully loaded. Therefore, it is necessary to increase the vehicle carrying rate to reduce transportation times and reduce the transportation cost when time permits. The reason why this situation still occurs is to meet urgent needs. Although the charging cost is high during the peak period, the prerequisite for reducing the cost requires priority to meet the demand for power exchange. Therefore, such a low single shipment would not be desirable from a cost standpoint, but such an arrangement is necessary from a demand standpoint.

Table 3 Optimal charging time and quantity distribution

Period of time	Electricity intensity	Electricity price	Number of batteries to start charging
7:00-11:00	Peak	1.234	0
11:00-19:00	Normal	0.856	267
19:00-23:00	Peak	1.234	74
23:00-7:00 (Next day)	Trough	0.376	247

Table 4 Optimal transportation time and quantity distribution of batteries

Period of time	Electricity intensity	Number of batteries shipped (BSS to BCS)	Number of batteries shipped (BCS to BSS)
7:00-11:00	Peak	0	0
11:00-19:00	Normal	366	260
19:00-23:00	Peak	40	1
23:00-7:00 (Next day)	Trough	0	314
7:00-11:00 (Next day)	Peak	0	6
11:00-13:30 (Next day)	Normal	0	0

4.3 Sensitivity analysis

The influence of demand distribution D on charging strategy

We did 100 rounds of sampling, and the sample size of each round was 61 to form 100 groups of samples. Fig. 5 and Fig. 6 show the distribution characteristics of 100 sets of samples.

We substituted 100 groups of random demand samples into the model to solve the optimal charging time and quantity distribution of batteries, as well as the transportation time and quantity distribution of batteries between BSS and BCS. The results show that the optimal charging time is basically concentrated in the 9th, 28th to 32nd and 33rd time particle size, and the number of batteries starting charging at the 9th and 33rd time particle size is the largest, with the number not less than 200. The number of batteries that started charging at time 28 to 32 was approximately 65. It can be seen that the uncertainty of demand has little influence on the final optimization result.

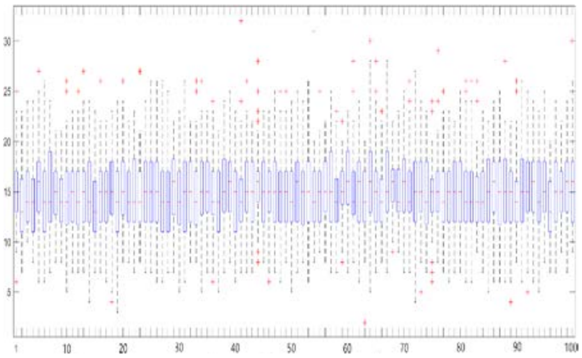


Fig. 5 100 sets of sample boxplot of electrical changing requirements

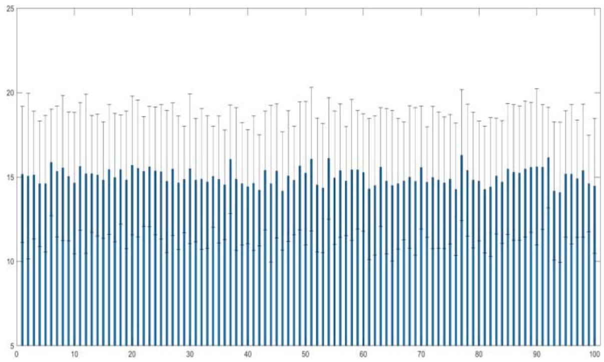


Fig. 6 100 sample mean and standard deviation distribution of electrical demand

The influence of initial battery storage ratio on charging strategy

The parameter represents the ratio of the available battery storage capacity in the BSS at the initial time. When the capacity of the BSS is determined, the value of represents the number of full batteries available in the initial BSS. It takes M (M = 10) time granularity for a depleted battery to be fully charged, which means that the system's electrical changing demands need to be met by artificially placed full batteries in the BSS during the initial 10 time granularity. Therefore, the value of has a very important impact on the operation of the system.

Fig. 7 and 8 show the value of $Q_{(t)}$ at each time granularity when N_{occup} takes values 0.5, 0.6, 0.7, 0.8, 0.9 and 1, respectively. When N_{occup} changes between 0.5 and 1, the model can concentrate most of the batteries on normal and low electricity prices. As the number of full batteries placed in the initial BSS increases, the number of batteries charged in the peak period decreases. This is because the system will give priority to meet the demand for electricity exchange. When the initial full battery quantity is low, the number of rechargeable batteries will increase at the peak.

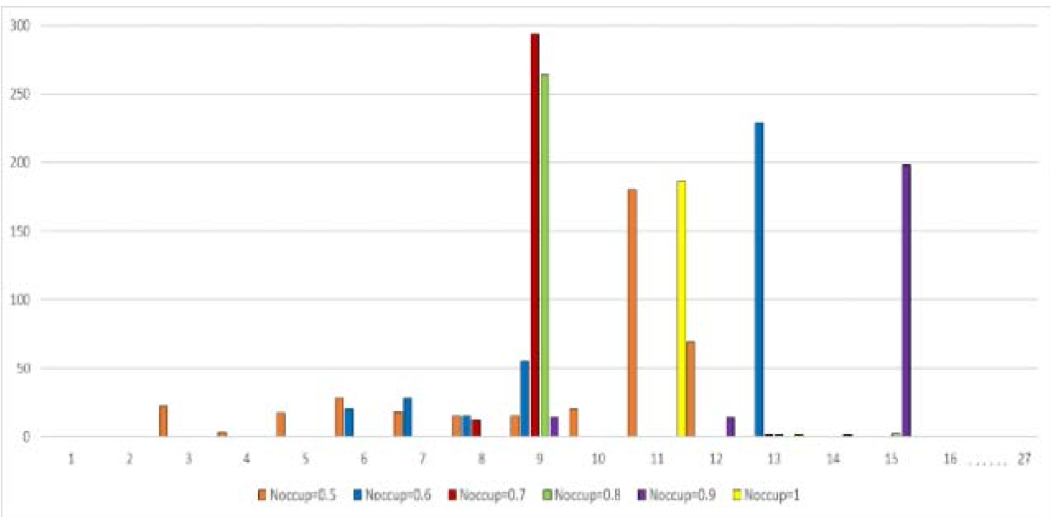


Fig. 7 The sensitivity change of N_{occup} to the optimal charging time

Fig. 9 is obtained by summing up $Q_{(t)}$ when N_{occup} takes different values according to the power consumption intensity.

Obviously, as the value of N_{occup} continues to increase, the total number of batteries charged during the research period decreases, and the number of batteries charged during peak periods also decreases.

As the value of the parameter N_{occup} continues to increase, the minimum cost value continues to decrease. This is because if the demand remains unchanged, the more fully charged batteries are artificially put into the system at the beginning, the fewer batteries need to be charged and transported in the later period of the system, thus reducing the cost. But in practice, the more batteries you put in, the more upfront costs you have. Table 6 shows that when N_{occup} takes 0.8, 0.9 and 1, the number of batteries charged in the peak period is not much different. Taking into account the trade-offs between the front and later costs, we recommend that the value of N_{occup} is 0.8.

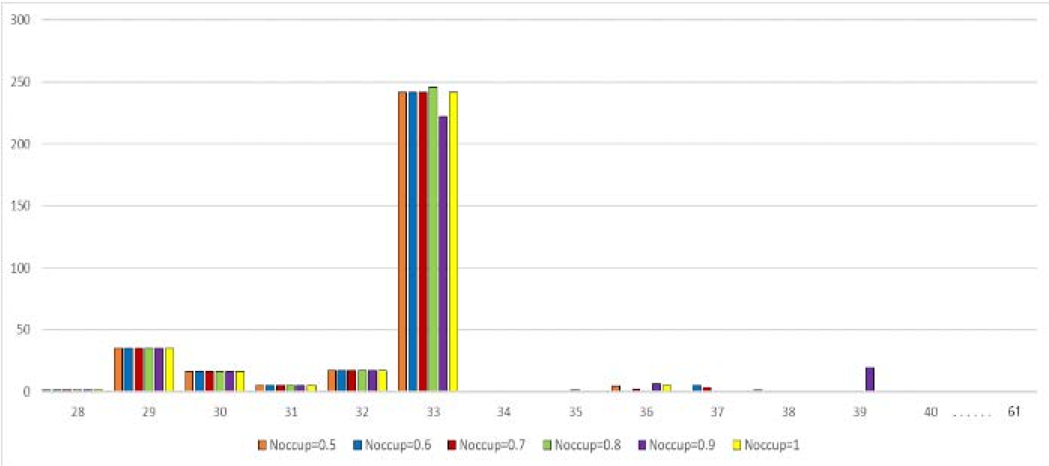


Fig. 8 The sensitivity change of N_{occup} to the optimal charging time

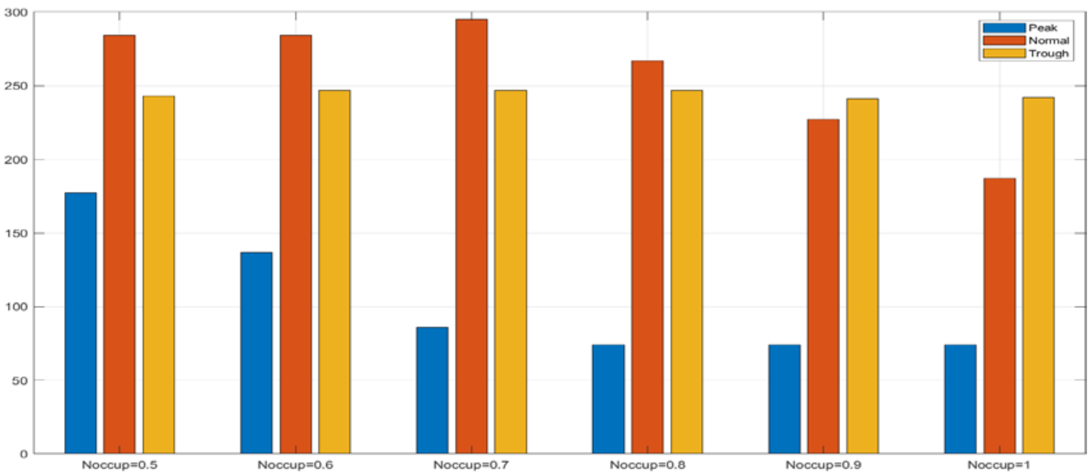


Fig. 9 Sensitivity analysis of N_{occup} to optimal charging time

Table 6 Distribution of optimal charging time and cost when takes different values

	Peak	Normal	Trough
N_{occup} = 0.5	177	284	243
N_{occup} = 0.6	137	284	247
N_{occup} = 0.7	86	295	247
N_{occup} = 0.8	74	267	247
N_{occup} = 0.9	74	227	241
N_{occup} = 1.0	74	187	242

5. Managerial implication

Based on the battery swapping mode, this study aims to introduce the joint optimization of battery charging and transportation and enrich the application of EVs in practice. Some constraints on the centralized battery charging and scheduling model are given special consideration. Although algorithm parameters and computational instances affect the calculation results, some specific conclusions are generalized as follows:

First, the model results are robust and easy to support the actual decision. In the example, the optimal solutions of 100 groups of random samples generated according to Poisson(15) are roughly the same, indicating that the random fluctuation of demand does not have a great impact on the final optimization results. This means that in the actual decision, even if there is some deviation in the estimation of future requirements, the optimization results still have high availability. What needs to be pointed out is that the above results are obtained on the premise that the electricity changing demands are subject to the Poisson(15) distribution. If the electricity changing demands are subject to the Poisson distribution with greater fluctuations or other random distributions, further tests are needed.

Second, the enterprises should focus on the time of centralized charging batteries to save charging costs and reduce power grid losses. From Table 3, nearly 90 % of the batteries can be charged in the off-peak period to avoid the peak period, so as to enjoy a discounted charging price. Furthermore, considering that the higher the power grid load is, the shorter the service life will be, our optimized charging scheme is helpful to reduce the peak load of the power grid and reduce the power grid loss, thus extending the power grid life. An interesting phenomenon is that about 10 % of the batteries are charged in the peak period, mainly because BSS must meet the demand of electrical changing in each period, but charging takes a certain time, so in some cases it must be charged in time regardless of the cost.

In order to save the charging cost in BSCS and improve the life of the power grid, when enterprises need to charge the depleted batteries, the optimization model should be adopted to centralized charge the rechargeable batteries in the period of low power consumption. Different enterprises have different requirements on service quality and response speed, so managers should choose a reasonable balance of charging time points according to their business status and demand network characteristics.

Third, the enterprises need to evaluate the tradeoffs between reducing charging costs and purchase cost of batteries at the initial time of BSS. From Table 9, if enterprises did not put enough full batteries in BSS at initial time, then in the later stage of BSCS operation, some batteries will be charged in the peak period in order to give priority to customer needs, which will increase the cost of battery charging. However, if enterprises placed more full batteries in the initial BSS, although in the later operation process there will be less batteries need to be charged and the charging cost will also be reduced, they need to pay more for the initial battery purchase cost. Considering that different enterprises have different service scope and volume, the configuration of parameter especially the purchase cost of EV batteries at initial time and battery charging costs should be well-balanced according to their business condition and demand network characteristics.

This paper makes full use of the dynamics of electricity prices to reduce the operating costs of enterprises and improve the security of the power grid. The battery swapping mode of electric vehicles has been tested and operated in many countries. This paper can provide relevant managers with a basis for decision-making. It contributes to the realization of the standardization of the electric vehicle battery industry and the optimal allocation of resources, and can also improve the satisfaction rate of demand, reduce charging costs, and accelerate the popularization and promotion of battery swapping mode.

6. Conclusion and the future work

In this study, the electrical changing behavior of EVs is decoupled into two processes of battery charging and battery exchange, and the problem is modeled as a mathematical model of transportation and charging cost minimization in the BSCS closed-loop supply chain to achieve reasonable battery scheduling. The validity of the model is proved by solving an example with LINGO. Finally, the results of the example are explained and analyzed, and the managerial implications of the centralized charging strategy and the optimal scheduling method of EV under the mode of electrical changing is emphasized.

The future work should further consider the benefit distribution among the three entities in BSCS under the cost optimization scheme. Transportation times between BSSs and BCS should also be further studied in the future. In addition, this study does not discuss the initial SOC of the battery in the battery swapping mode. In fact, the battery capacity for the battery exchange has a certain randomness. Future research can divide the battery initial SOC into segments to further refine the battery charging time.

Acknowledgement

This work is supported by Beijing Social Science Fund (Grant no. B19SK00630), Beijing Intelligent Logistics System Collaborative Innovation Center (Grant no. BILSCIC-2019KF-24), and Beijing Logistics Informatics Research Base.

References

[1] Jie, W., Yang, J., Zhang, M., Huang, Y. (2019). The two-echelon capacitated electric vehicle routing problem with battery swapping stations: Formulation and efficient methodology, *European Journal of Operational Research,* Vol. 272, No. 3, 879-904, doi: 10.1016/j.ejor.2018.07.002.

[2] Su, W., Eichi, H., Zeng, W., Chow, M.-Y. (2012). A survey on the electrification of transportation in a smart grid environment, *IEEE Transactions on Industrial Informatics*, Vol. 8, No. 1, 1-10, doi: 10.1109/TII.2011.2172454.

[3] Gong, D., Tang, M., Liu, S., Xue, G., Wang, L. (2019). Achieving sustainable transport through resource scheduling: A case study for electric vehicle charging stations, *Advances in Production Engineering & Management*, Vol. 14, No. 1, 65-79, doi: 10.14743/apem2019.1.312.

[4] Barmaki, R., Ilkhani, M., Salehpour, S. (2016). Investigation of energy usage and emissions on plug-in and hybrid electric vehicle, *Tehnički Vjesnik – Technical Gazette,* Vol. 23, No. 3, 899-906, doi: 10.17559/TV-20140928112417.

[5] Pirbadali-Somarin, A., Peyghambarzadeh, S.M. (2020). Air pollution by heavy metals from petrochemical incinerators: Measurement and dispersion modelling, *Environmental Engineering and Management Journal*, Vol. 19, No. 3, 379-390, doi: 10.30638/eemj.2020.036.

[6] Wang, G., Xu, Z., Wen, F., Wong, K.P. (2013). Traffic-constrained multiobjective planning of electric-vehicle charging stations, *IEEE Transactions on Power Delivery*, Vol. 28, No. 4, 2363-2372, doi: 10.1109/TPWRD.2013.2269142.

[7] Wang, G., Zhao, J., Wen, F., Xue, Y., Ledwich, G. (2015). Dispatch strategy of PHEVs to mitigate selected patterns of seasonally varying outputs from renewable generation, *IEEE Transactions on Smart Grid,* Vol. 6, No. 2, 627-639, doi: 10.1109/TSG.2014.2364235.

[8] Tao, Y., Huang, M., Chen, Y., Yang, L. (2020). Orderly charging strategy of battery electric vehicle driven by real-world driving data, *Energy,* Vol. 193, Article No. 116806, doi: 10.1016/j.energy.2019.116806.

[9] Pieltain Fernández, L., Gómez San Román, T., Cossent, R., Mateo Domingo, C., Frías, P. (2011). Assessment of the impact of plug-in electric vehicles on distribution networks, *IEEE Transactions on Power Systems,* Vol. 26, No. 1, 206-213, doi: 10.1109/TPWRS.2010.2049133.

[10] Clement-Nyns, K., Haesen, E., Driesen, J. (2010). The impact of charging plug-in hybrid electric vehicles on a residential distribution grid, *IEEE Transactions on Power Systems,* Vol. 25, No. 1, 371-380, doi: 10.1109/TPWRS.2009.2036481.

[11] Zhang, X., Wang, G. (2016). Optimal dispatch of electric vehicle batteries between battery swapping stations and charging stations, In: *Proceedings of 2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, Massachusetts, USA, 1-5, doi: 10.1109/PESGM.2016.7741893.

[12] Tang, M., Gong, D., Liu, S., Lu, X. (2017). Finding key factors affecting the locations of electric vehicle charging stations: A simulation and ANOVA approach, *International Journal of Simulation Modelling*, Vol. 16, No. 3, 541-554, doi: 10.2507/IJSIMM16(3)CO15.

[13] Stüdli, S., Griggs, W., Crisostomi, E., Shorten, R. (2014). On optimality criteria for reverse charging of electric vehicles, *IEEE Transactions on Intelligent Transportation Systems,* Vol. 15, No. 1, 451-456, doi: 10.1109/TITS.2013.2271953.

[14] Deilami, S., Masoum, A.S., Moses, P.S., Masoum, M.A.S. (2011). Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile, *IEEE Transactions on Smart Grid*, Vol. 2, No. 3, 456-467, doi: 10.1109/TSG.2011.2159816.

[15] Qiu, D., Yang, H., Lai, M. (2018). Optimal bidding strategy and EV charging management of intelligent regional aggregators based on CVaR method, *Systems Engineering – Theory and Practice*, Vol. 38, No. 8, 1994-2005, doi: 10.12011/1000-6788(2018)08-1994-12.

[16] Zheng, Y., Dong, Z.Y., Xu, Y., Meng, K., Zhao, J.H., Qiu, J. (2014). Electric vehicle battery charging/swap stations in distribution systems: Comparison study and optimal planning, *IEEE Transactions on Power Systems*, Vol. 29, No. 1, 221-229, doi: 10.1109/TPWRS.2013.2278852.

[17] Kang, Q., Wang, J.-B., Zhou, M.-C., Ammari, A.C. (2016). Centralized charging strategy and scheduling algorithm for electric vehicles under a battery swapping scenario, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 3, 659-669, doi: 10.1109/TITS.2015.2487323.

[18] Zheng, J., Xie, T., Liu, F., Wang, W., Du, P., Han, Y. (2017). Electric vehicle battery swapping station coordinated charging dispatch method based on CS algorithm, In: *Proceedings of 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC),* Chongqing, China, 150-154, doi: 10.1109/ITOEC.2017.8122400.

[19] Yang, S., Yao, J., Kang, T., Zhu, X. (2014). Dynamic operation model of the battery swapping station for EV (electric vehicle) in electricity market, *Energy*, Vol. 65, 544-549, doi: 10.1016/j.energy.2013.11.010.

[20] Song, Y., Li, J., Ji, G., Xue, Z. (2016). Study on the typical mode of EV charging and battery swap infrastructure interconnecting to power grid, In: *Proceedings of 2016 China International Conference on Electricity Distribution (CICED),* Xi'an, China, 1-4, doi: 10.1109/CICED.2016.7576066.

[21] Rao, R., Zhang, X., Xie, J., Ju, L. (2015). Optimizing electric vehicle users' charging behavior in battery swapping mode, *Applied Energy*, Vol. 155, 547-559, doi: 10.1016/j.apenergy.2015.05.125.

[22] Wang, T., Ling, Y., Liu, J., Han, J., Chen, L., Zhao, H., Feng, S., Han, T. (2018). Integrated optimal configuration of electric vehicle charging and battery-swapping station based on ordered charging strategy, In: *Proceedings of 2018 China International Conference on Electricity Distribution (CICED)*, Tianjin, China, 2237-2241, doi: 10.1109/CICED.2018.8592420.

[23] Infante, W.F., Ma, J., Chi, Y. (2016). Operational strategy and load profile sensitivity analysis for an electric vehicle battery swapping station, In: *Proceedings of 2016 IEEE International Conference on Power System Technology (POWERCON),* Wollongong, New South Wales, Australia, 1-6, doi: 10.1109/POWERCON.2016.7753877.

[24] Sarker, M.R., Pandžić, H., Ortega-Vazquez, M.A. (2015). Optimal operation and services scheduling for an electric vehicle battery swapping station, *IEEE Transactions on Power Systems,* Vol. 30, No. 2, 901-910, doi: 10.1109/TPWRS.2014.2331560.

[25] Wang, Y., Yang, Y., Zhang, N., Huang, M. (2017). An integrated optimization model of charging station/battery-swap station/energy storage system considering uncertainty, In: *Proceedings of 2017 IEEE International Conference on Energy Internet (ICEI),* Beijing, China, 77-82, doi: 10.1109/ICEI.2017.21.

Appendix

Table 7 Abbreviations of variables

Abbreviations	Meaning
EV	Electric vehicle
BCS	Battery charging station
BSS	Battery swapping station
BSCS	Battery swapping and charging system
SOC	SOC is the state of charge, which is used to reflect the remaining capacity of the battery
DB	The depleted battery
FB	The fully charged battery

APEM journal
Advances in Production Engineering & Management **ISSN 1854-6250**
Volume 16 | Number 3 | September 2021 | pp 372–384 Journal home: apem-journal.org
https://doi.org/10.14743/apem2021.3.407 Original scientific paper

A multi-objective selective maintenance optimization method for series-parallel systems using NSGA-III and NSGA-II evolutionary algorithms

Xu, E.B.[*,a], Yang, M.S.[a], Li, Y.[a], Gao, X.Q.[a], Wang, Z.Y.[a], Ren, L.J.[a]

[a]School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an, P.R. China

A B S T R A C T

Aiming at the problem that the downtime is simply assumed to be constant and the limited resources are not considered in the current selective maintenance of the series-parallel system, a three-objective selective maintenance model for the series-parallel system is established to minimize the maintenance cost, maximize the probability of completing the next task and minimize the downtime. The maintenance decision-making model and personnel allocation model are combined to make decisions on the optimal length of each equipment's rest period, the equipment to be maintained during the rest period and the maintenance level. For the multi-objective model established, the NSGA-III algorithm is designed to solve the model. Comparing with the NSGA-II algorithm that only considers the first two objectives, it is verified that the designed multi-objective model can effectively reduce the downtime of the system.

A R T I C L E I N F O

Keywords:
Maintenance;
Series-parallel system;
Maintenance decision model;
Multi-objective optimization;
Selective maintenance;
Evolutionary algorithms;
Non-dominated sorting genetic algorithm;
NSGA-II;
NSGA-III

**Corresponding author:*
1180211005@stu.xaut.edu.cn
(Xu, E.B.)

Article history:
Received 1 July 2021
Revised 22 October 2021
Accepted 25 October 2021

1. Introduction

The safe and reliable operation of equipment is the primary condition for enterprises to ensure production efficiency. Reasonable maintenance methods can effectively guarantee the reliable operation of equipment and reduce the maintenance cost of enterprises. Effective, reliable and economical equipment maintenance plays an increasingly important role in enterprise production and operation [1]. In the manufacturing environment, manufacturing systems often contain multiple devices, forming a multi-device system, and there are often complex dependencies between them. How to make reasonable maintenance and maintenance decisions is the research focus of predictive maintenance. Common maintenance decision-making methods include mathematical model analysis method, Markov decision model method, simulation model method, etc. [2].

Mathematical analytic method is generally based on the knowledge in the field of operations research to study the application and planning of various maintenance resources, establish the corresponding model and solve it. Rashidnejad *et al.* proposed a dual-objective model and solution to the geographically dispersed asset maintenance plan based on NSGA-II (Non-Dominated Sorting Genetic Algorithm II) algorithm, and evaluated the effectiveness and performance of the proposed model through the actual case of bank ATM (Automatic Teller Machine) maintenance [3]; Pandey *et al.* made decisions about the selection of equipment to be maintained and the level of maintenance to be performed by the maintenance equipment during the fixed downtime of the continuous generation system [4]; Dao *et al.* established a maintenance optimization model of multi-state serial-parallel system based on resource availability and component maintenance time and cost dependence, and used genetic algorithm to solve the optimization problem [5]. Khatab and Khatab assumed that the duration of the next task and the duration of rest were random variables with known distribution, and modeled the final selective maintenance optimization problem as a mixed integer nonlinear stochastic programming [6-7]; Chaabane proposed a selective maintenance planning (SMP) model to optimize maintenance and allocation decisions in systems running multiple tasks[8]; Cheng *et al.* conducted joint modeling and Optimization on output, product quality and predictive maintenance of series parallel multi-stage production system, and reasonably allocated maintenance resources according to the importance of equipment [9]; Lu *et al.* optimized the preventive maintenance problem of multi equipment series production system with buffer zone by using genetic algorithm [10]; For the Series production system without buffer, Leng *et al.* made decisions and optimized the production lot size and the imperfect preventive maintenance of the equipment under the conditions of both shortage and inventory [11]. According to the resource demand priority of each process in the production system, Li *et al.* established the opportunity maintenance strategy of multi-resource constrained serial-parallel system. Apart from the above example, a different structure of manuscript may be accepted if it is the most suitable and effective style for the contents of the manuscript [12].

Markov decision model is a mathematical model to simulate the Random Strategy and Return of agents in the environment, and the state of environment has Markov property, which is widely used in dynamic programming, random sampling, decision optimization and other fields. Bousdekis *et al.* proposed a main event-driven model for joint maintenance and logistics optimization in the Industrial Internet environment, which combines the Markov decision process (MDP) and embeds the model into the event driven information system to make maintenance and logistics optimization decisions [13]; Gerum *et al.* proposed a new method to predict rail defects, and determined the best maintenance strategy through Semi-Markov decision process model [14]. Xu *et al.* Proposed a Dynamic state maintenance decision model based on Markov, which was used to make maintenance decisions for the Major components of equipment [15].

Pei *et al.* constructed an imperfect maintenance stochastic degradation model based on the Wiener process, and established a maintenance decision model to decide variables by the inspection intervals and maintenance threshold, and solved the model through numerical simulation [16]; Nguyen *et al.* proposed a predictive maintenance strategy with multi-level decision, which considers respectively system level and component level maintenance decision-making process, and uses Monte Carlo simulation technique to evaluate the maintenance cost rate [17]; Dao and Zuo studied the selective maintenance problem of a multi-state series system working under variable load conditions in the next task, and simulated the degradation of multi-state components through Monte Carlo simulation methods and evaluated the system reliability to determine the best choice maintenance strategy to maximize the expected reliability of the system for the next task within the range of available resources [18].

Studies have shown that the shutdown time is usually assumed to be constant, most scholars only consider the two optimization objectives of minimizing maintenance cost and maximizing the probability of completing the next task. However, in the enterprise continuous production line, the maintenance decision-making model and personnel allocation model are not considered together, when making maintenance decisions for the common series-parallel system of multiple devices. In each decision-making cycle, it is not only to solve the problems of which equipment needs to be maintained, the maintenance level of each equipment and the task assignment, but also to obtain the optimal downtime of the system. Therefore, we take the minimizing downtime as the third objective of the decision scheduling model. Then, we design a NSGA-III algorithm to solve the three-objective model. Compared with the NSGA-II algorithm of two objectives, the three objective decision model can achieve optimal downtime of system maintenance and detailed dispatch of maintenance tasks, and effectively reduce the downtime of the system.

2. Description of selective maintenance problem for series-parallel system

In the continuous production process, the production line is generally composed of multiple devices in series and parallel. In the multi equipment system, there are complex relationships among the equipment, such as economic dependence, structural dependence, random fault dependence and resource dependence. Economic dependence refers to the maintenance or inspection of multiple equipment at the same time under the condition of limited budget. This moment, there is an economic dependence among the equipment. Random dependence means that the deterioration process or failure time of various equipment has random correlation to some extent. Structural dependence refers to that the failure of one equipment may lead to the deterioration or failure of other equipment, and the maintenance of one component in the unit also means the maintenance of other components. Resource dependence refers to that the maintenance personnel are responsible for the maintenance activities of various units or systems, the limited spare parts inventory is used to replace multiple equipment, or selectively maintain the multi-equipment system in a limited time window.

Therefore, it is necessary to consider various dependence relationships between equipment in maintenance decision-making. The structure of series-parallel system is shown in Fig. 1. Series-parallel system usually performs continuous production tasks. Selective maintenance of equipment in the time interval between two continuous tasks can improve the reliability of the system when new tasks start. Due to the limited resources to complete maintenance activities, it is particularly important to determine the maintenance strategy based on the system requirements.

In the maintenance of series-parallel system, not only the loss cost of each equipment, but also the probability of the overall failure system should be considered. Therefore, it is necessary to determine the system shutdown time and the length of shutdown time during the maintenance of series parallel system, because the shorter the shutdown time, the smaller the impact on production. In order to solve the problem that traditional maintenance decision-making model of series-parallel system does not consider task dispatch and the system downtime is usually assumed to be constant, a multi-objective maintenance decision-making model of integrated maintenance assignment system is established by taking the shortest system downtime as one of the optimization objectives, and genetic algorithm is designed to solve the model.

Suppose that a production system is composed of n independent subsystems in series, each subsystem i is composed of m_i independent identical subsystems in parallel. Each device in the system can be expressed as E_{ij}, where i is the sub-system index and j is the sub-device index.

Each device in the system has two possible states, denoted by 0 (Complete failure state) and 1 (normal operation state). Since the whole system is composed of several subsystems in series, the whole system can work when all subsystems are running normally, and the device parallel structure in the subsystem, at least one device can work when they are all in normal operation.

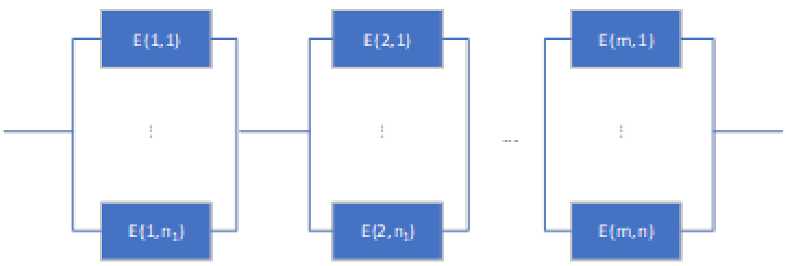


Fig. 1 Series parallel system structure

Within a given time period, the system needs to perform a series of consecutive identical tasks, and maintenance activities are performed only during the rest time between adjacent tasks. In order to facilitate future discussion, it is assumed that the amount of maintenance resources (budget, time, etc.) required for a maintenance activity is determined, and the reliability of components, subsystems or systems is given in advance, which is the possibility of successful completion of a given task.

Question hypothesis:

1. The system is composed of multiple independent repairable binary state devices, that is, the device is in a normal working state or a failure shutdown state. The life of each device in the initial state is different, and the life distribution parameters are known.
2. Maintenance activities (imperfect maintenance, preventive replacement, corrective maintenance) are performed at most once for each equipment during each rest cycle, in which corrective maintenance and preventive replacement are both perfect maintenances.
3. System components will not be aging during rest, that is, the life of components depends on the operation.
4. The cost of corrective maintenance is higher than that of preventive maintenance, and the mean time for corrective maintenance on each device is longer than the average time for preventive maintenance.
5. During the mission, the system does not perform any maintenance activities. When the faulty components are repaired to a minimum degree, the failure rate remains unchanged.
6. When necessary, all essential maintenance resources (maintenance personnel, tools and rest hours) are available. If the maintenance resource constraints are exceeded, the corresponding parameters should be punished, and the model is still holding.

3. Modeling of series-parallel maintenance system

In order to establish the selective maintenance model of series-parallel system, the following symbols are defined:

- A_{ij} is the effective service age of each equipment at the beginning of maintenance period;
- B_{ij} is the effective service age of each equipment at the ending of maintenance period;
- Y_{ij} is the state of each equipment at the ending of maintenance period;
- X_{ij} is the state of each equipment at the beginning of maintenance period (if the equipment works normally at the beginning of maintenance $X_{ij} = 1$, otherwise it is 0. if the equipment works normally when the maintenance ends $Y_{ij} = 1$, otherwise it is 0);
- l_{ij} is the level of maintenance performed by the equipment in the current rest cycle (0 indicates that no maintenance is performed, 1 indicates preventive imperfect maintenance, 2 indicates preventive substitution, and 3 indicates corrective maintenance);
- K is the number of maintenance personnel;
- k is the index of maintenance personnel;
- $P_{ij,t}$ is the cumulative failure probability of equipment at time, which is given by the prediction model;
- DT_{ij} is the expected downtime for unexpected equipment failure;
- δ_{ij} is the unit downtime cost for equipment;
- T_{IJkL} is the time it takes for maintenance personnel k to perform maintenance level l on equipment E_{ij};
- C_{IJkL} is the cost it takes for maintenance personnel k to perform maintenance level l on equipment E_{ij};
- when equipment E_{ij} is maintained at level l by maintenance personnel k, w_{ijlk} is 1, otherwise it is 0;
- T_{stop} is the length of downtime for maintenance;
- $f(t)$ is the downtime cost of the system over time;

Assuming that the life of each device E_{ij} in system obeys Weibull distribution, and the distribution of its shape parameters and scale parameters are β_{ij} and α_{ij}, then the reliability $R_{ij}(t)$ of equipment E_{ij} is:

$$R_{ij}(t) = \exp\left(-\left(\frac{t}{\eta_{ij}}\right)^{\beta_{ij}}\right) \tag{1}$$

The total reliability of the whole system at time t is:

$$\Re(t) = \prod_{i=1}^{m} R_i(t) = \prod_{i=1}^{m}\left(1 - \prod_{j=1}^{n}(1 - R_{ij}(t))\right) \tag{2}$$

The probability that device E_{ij} can still run normally after performing the next task with length u is as follows:

$$R_{ij}^{c} = \frac{R_{ij}(A_{ij} + U)}{R_{ij}(A_{ij})} \tag{3}$$

The reliability of the whole system to successfully complete the next task with length u is as follows:

$$\Re^{c} = \prod_{i=1}^{m}\left(1 - \prod_{j=1}^{n}(1 - R_{ij}^{c})\right) \tag{4}$$

The cost of maintenance:

$$C_{mt} = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{K}\sum_{l=0}^{L_{ij}} C_{IJkL} \cdot w_{tijlk} \tag{5}$$

The maintenance personnel cost:

$$C_{hire} = K \cdot c_p \cdot T_{stop} \tag{6}$$

The maintenance time:

$$T_{stop} = max\left(\sum_{i=1}^{m} \cdot \sum_{j=1}^{n} \cdot \sum_{l=1}^{L_{ij}} t_{ijkl} \cdot w_{tijlk}\right) \quad k \in K \tag{7}$$

Objective function 1: Minimize maintenance costs

$$\text{Min } C_{total} = C_{mt} + C_{hire} \tag{8}$$

Objective function 2: Maximize the probability of the system successfully executing the next task

$$\text{Max } \Re^{c} = \prod_{i=1}^{m}\left(1 - \prod_{j=1}^{n}(1 - R_{ij}^{c})\right) \tag{9}$$

Objective function 3: Minimize maintenance time

$$\text{Min } T_{stop} \tag{10}$$

Constraint condition:

- Maintenance activities are performed at most once for each equipment during shutdown:

$$\sum_{k}^{K}\sum_{l=0}^{L_{ij}} w_{ijkl} \le 1, \quad \forall i, j \tag{11}$$

- Maintenance duration cannot be greater than system downtime:

$$\sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{l=0}^{L_{ij}} t_{ijkl}\cdot w_{ijkl} \leq T_{stop}, \quad \forall k \tag{12}$$

- If the equipment is faulty at the start of the rest, correction must be performed:

$$\sum_{k}^{K}\sum_{l=0}^{L_{ij}} w_{ijkl} = 3\,, \qquad X_{ij} = 0, \forall i,j \tag{13}$$

- The constraint of Maintenance decision and failure probability:

$$w_{ijkl} \in \{0,1\}, 0 \leq R_{ij}(t) \leq 1, \quad \forall i,j,k,l,t \tag{14}$$

4. Designing of multi-objective genetic algorithm

The traditional genetic algorithm is only suitable for the single-objective optimization problem, and cannot solve the optimization model similar to the selective maintenance of series-parallel system.

Srinivas and Deb proposed a Non-Dominated Sorting Genetic Algorithm (NSGA), using the characteristics of genetic algorithm parallel batch to obtain as many Pareto solutions as possible, but in the later use process there are high computational complexity, lack of elite and need to share parameters [19]. Therefore, Deb *et al.* proposed a Non-Dominated Sorting Genetic Algorithm II, which solved the above problems of NSGA, and could find the non-dominated solution set with good convergence and dispersion in the two-objective optimization problem [20].

Zhao *et al.* [21] proposed a time-dependent and bi-objective vehicle routing problem with time windows (TD-BO-VRPTW). The non-dominated sorting genetic algorithm II (NSGA-II) is adopted to obtain the Pareto optimal solution set. Through comparing these results with solutions in the Pareto front, the results in Pareto front are competitive because there is a trade-off between two objectives. Liu and Zhang [22] established a multi-objective planning model. This model can solve the dual uncertainty demand problems of number and delivery time when orders are emergent or are modified for equipment manufacturing enterprises. The NSGA-II genetic algorithm is used to solve the model. Targeting at the problems existing in the multi-objective scheduling of traditional flexible job shop and the complexity of multi-resource allocation, Zhong *et al.* [23] established an improved calculation model considering the optimization of such four targets as completion time, labour distribution, equipment compliance and production cost. The multi-objective integrated constraint optimization algorithm was designed and the Pareto solution set following different rules based on the NSGA-Pi algorithm was finally obtained.

However, when the optimization objectives increase, the multi-objective optimization algorithm encounters new challenges: 1) the proportion of non-dominated solutions in the target vector set increases exponentially; 2) the calculation quantity of diversity protection operator is huge;3) surface visualization is difficult to achieve. In order to solve the above problems, Deb *et al.* proposed a multi-objective algorithm based on the framework of NSGA-II, which is called NSGA-III [20]. This algorithm has obtained good results in solving the multi-objective optimization problem with 3-15 objectives. Therefore, we select the three objectives in the NSGA-III algorithm to solve the multi-objective selective maintenance problem.

The NSGA-III algorithm flow is shown in Fig. 2. In essence, NSGA-III adopts a similar framework of NSGA-II, the difference is mainly the change of selection mechanism. NSGA-II mainly uses crowding to sort, its effect in high-dimensional target space is not obvious, and NSGA-III maintain population diversity by introducing widely distributed reference points.

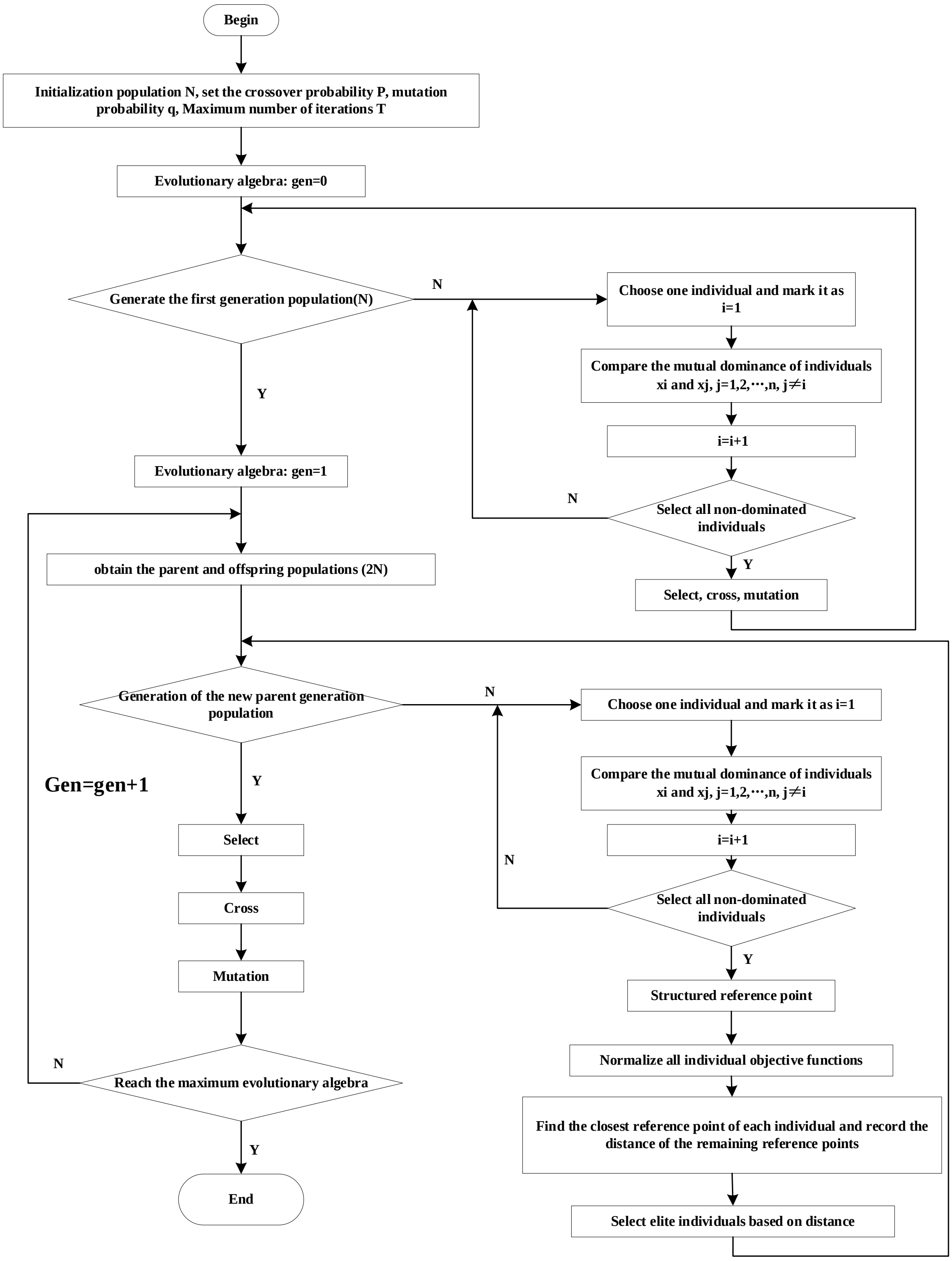


Fig. 2 NSGA-III algorithm flow

Aiming at the multi-objective optimization model for series-parallel system, the coding, crossover and mutation operators of NSGA-III algorithm are designed as follows.

4.1 Encoding

For the multi-objective optimization problem of selective maintenance in series-parallel system, each chromosome needs to determine three problems: 1) the equipment that needs to perform maintenance operation during rest period; 2) The level of maintenance operations performed by the equipment to be maintained; 3) Maintenance task assignment issues, so three issues need to be covered when encoding. Assuming there are *N* devices in the system, the chromosome is represented by a matrix of *2N* rows. Among them, the first *N* columns of the chromosome correspond to the equipment in the system, and the coding of the corresponding position represents the maintenance operation level that needs to be performed during the current rest period (0 indicates that no maintenance is performed, 1 indicates preventive imperfect maintenance, 2 indicates preventive substitution, and 3 indicates corrective maintenance). The latter N is the distribution of the corresponding equipment maintenance tasks. If the equipment i decides not to perform any maintenance operation, the corresponding position is 0. If the maintenance personnel 2 decides to perform preventive replacement, the corresponding position is 2, indicating that the maintenance operation is performed by the maintenance personnel 2. Therefore, Fig. 3 shows an example of chromosome encoding.

Fig. 3 Encoding

4.2 Crossover operator

The crossover operation adopts the double-cut point crossover method. First, two positions are generated in the chromosome part that determines the maintenance level of the equipment. Then, the fragments of the two chromosome crossover intervals are exchanged, and the corresponding task assignment gene fragments are exchanged in the same time. This crossover method ensures the feasibility of generating new offspring. The crossover operation is shown in Fig. 4.

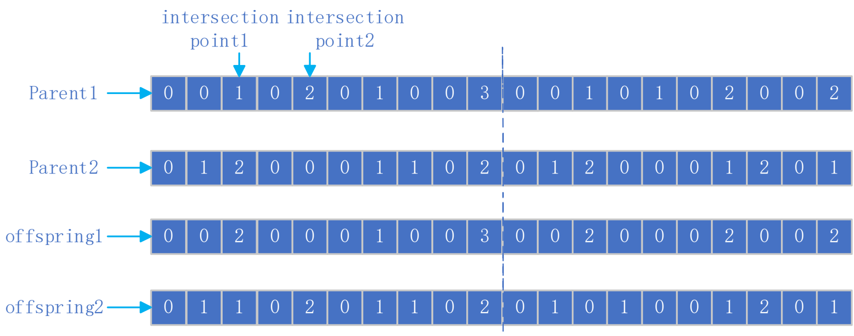


Fig. 4 Crossover operator

4.3 Mutation operator

In this paper, two types of mutation operators are used, and one is randomly selected from the two mutation operators for mutation operation each time. The first mutation operator performs mutation operation on the chromosome part which determines the equipment maintenance level, and the second mutation operator performs mutation operation on the chromosome part of equipment maintenance allocation.

The first mutation operator generates two arbitrary points in the first half of the chromosome, cross the encoding of the corresponding position, and exchange the encoding of the corresponding dispatching chromosome region, as shown in Fig. 5.

The second mutation operator randomly generates two non-zero gene positions in the second half of the chromosome, and then cross codes the corresponding positions, as shown in Fig. 6.

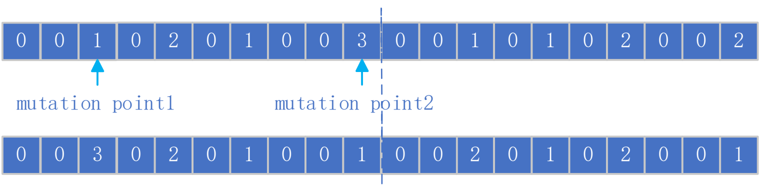


Fig. 5 Mutation operator 1

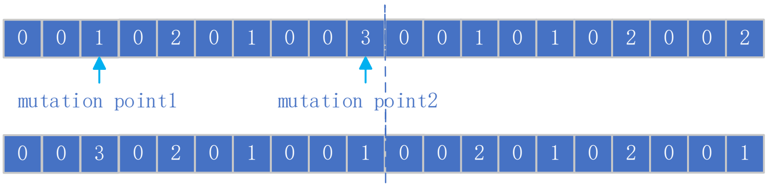


Fig. 6 Mutation operator 2

The above two mutation methods ensure the feasibility of chromosome after mutation, so it is not necessary to verify the rationality of coding.

5. A case study and a comparative analysis of NSGA-II and NSGA-III

The equipment composition of a production line is shown in Fig. 7, which is consisted of five subsystems composed of 14 components. The Weibull distribution life of each equipment and the shape and scale parameters represented have been given. PM represents the implementation of preventive replacement, IPM represents the implementation of imperfect maintenance, and CM represents the implementation of corrective maintenance. In addition, other parameters are given in Table1, in which the units of various maintenance costs are all yuan, and the unit of maintenance duration is day. Suppose the downtime cost of the system is RMB500 per day and the hiring cost of a single maintenance worker is RMB50 per day.

In order to verify the validity of the model in this paper, NSGA-II algorithm and NAGA-III algorithm are used to solve the model respectively. The goal of NSGA-II is to minimize maintenance costs and maximize the success rate of the next task, the goal of NAGA-III algorithm increases the minimum downtime maintenance time on the basis of NSGA-II. In both algorithms, the population is 500, the crossover probability is 0.8, the mutation probability is 0.1, and the number of iterations is 500.

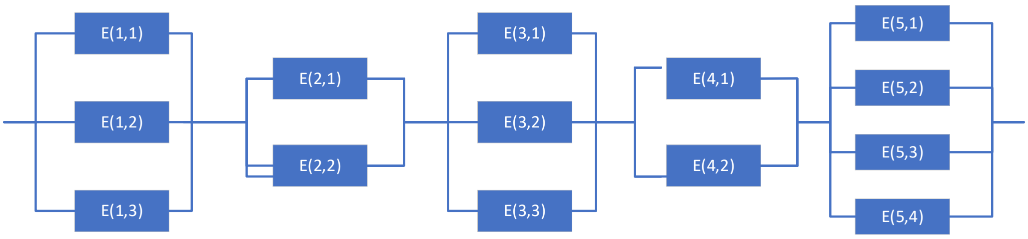


Fig. 7 System structure

Table 1 Example parameter table

Serial number	α	β	PM cost	PM time	CM cost	CM time	Operating time	IPM cost	IPM time
E11	1.5	250	476	1	952	1.5	110	158	0.5
E12	2.4	380	653	3	1306	4.5	150	217	1
E13	1.6	280	962	2	1924	3	170	320	1
E21	2.5	400	323	2	646	3	120	107	1
E22	1.5	280	185	3	370	4.5	180	61	1
E31	2.4	340	639	2	1278	3	100	213	1
E32	2.5	260	812	2	1624	3	130	270	1
E33	2	280	391	3	782	4.5	170	130	1
E41	1.2	260	672	1	1344	1.5	150	224	0.5
E42	1.4	350	188	1	376	1.5	120	62	0.5
E51	2.8	400	294	3	588	4.5	180	98	1
E52	1.5	350	297	1	594	1.5	130	99	0.5
E53	2.4	300	394	2	788	3	100	131	1
E54	2.2	450	712	1	1424	1.5	150	237	0.5

The running time of NSGA-II algorithm is 460 s, and the number of Pareto solutions is 500. The calculation results are shown in Fig. 8. The surface formed by the optimal set in space is called Pareto fronts. All solutions in Pareto front are not dominated by other solutions outside or within the Pareto front curve. Therefore, these non-dominated solutions have the least goal conflict than other solutions, which can provide a better choice space for decision makers. So, after calculation, the number of Pareto front solutions is 26.

The running time of NSGA-III algorithm is 474 s, and the number of Pareto solutions obtained is 496. The results of NSGA-III algorithm are shown in Fig. 9. After calculation, the number of Pareto front solutions is 19.

The Pareto optimal solution set obtained by the algorithm only provides the non-inferior solution of the problem to three objectives, and there is no single objective optimal. Therefore, it is necessary to select according to the expected degree of each objective of the decision maker. Assuming that the decision maker is most concerned about the probability of the system successfully completing the next phase of the task and the maintenance cost, the chromosome with the smallest maintenance cost can be selected from the solution set with the reliability of completing the next task greater than 0.85.

The maintenance scheduling decision scheme solved by NSGA-II algorithm is shown in Fig. 10. The equipment that performs maintenance is selected as E11, E12, E22, E31, E33 and E42. The maintenance level is preventive replacement, and the final maintenance cost is 3182. The probability of the system successfully completing the next task is 0.853. The shutdown maintenance time of the system is 5 days, and the number of maintenance personnel enabled is 1.

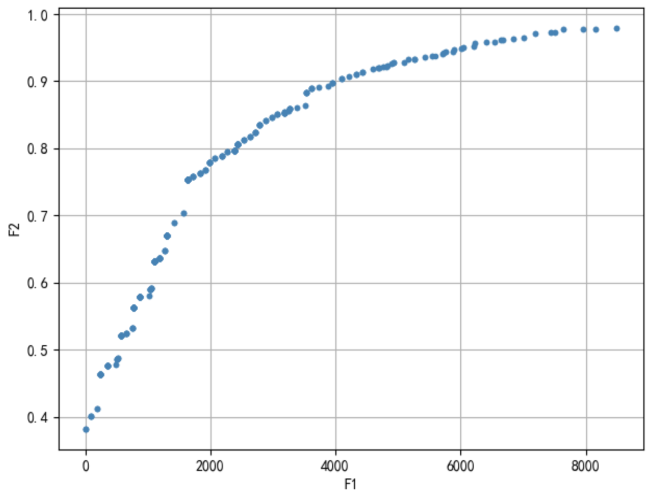


Fig. 8 NSGA-II algorithm results

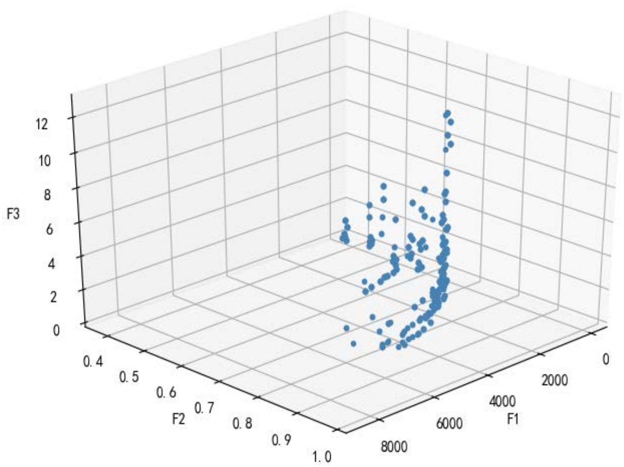


Fig. 9 NSGA-III algorithm results

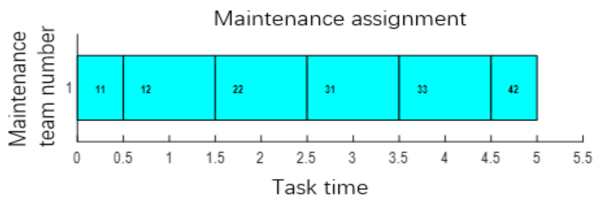


Fig. 10 NSGA-II algorithm maintenance scheduling scheme

The maintenance scheduling decision scheme solved by NSGA-III algorithm is shown in Fig. 11. The equipment that performs maintenance is selected as E11, E12, E22, E31, E33, E42, and E52. The machine E33 performs imperfect maintenance, and the other equipment performs preventive replacement maintenance. The final maintenance cost is 3168. The probability of the system successfully completing the next task is 0.850. The system shutdown maintenance time is 2 days, and the number of maintenance personnel enabled is 3.

The comparison of the two algorithms is shown in Fig. 12. It can be seen that when the maintenance cost is close to the probability of completing the next task, the downtime of the maintenance scheduling result solved by NSGA-III algorithm is less than that of NSGA-II algorithm.

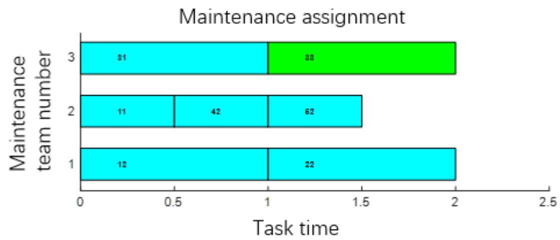


Fig. 11 NSGA-III algorithm maintenance scheduling scheme

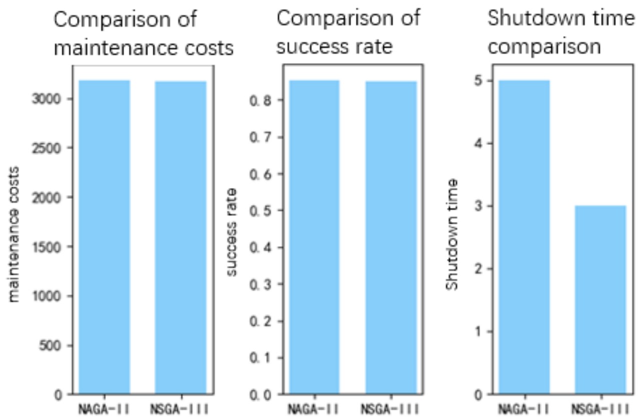


Fig. 12 Comparison of two algorithms

6. Conclusion

Aiming at the problems of selective maintenance for series-parallel systems, including simply assuming downtime and not considering resource maintenance dispatch, firstly, we establish a multi-objective selective maintenance model with the purpose of minimizing downtime, minimizing maintenance cost and maximizing the probability of completing the next task. Then, the genetic algorithm is selected to solve the model, and the coding method, crossover operator and mutation operator are designed in detail. Finally, in order to verify the effectiveness of the designed strategy, we design NSGA-II and NSGA-III algorithms with specific examples. These two algorithms are respectively used to solve the two objective maintenance decision-making model which only considers minimizing maintenance and maximizing the probability of the system completing the next task, and the three objective maintenance decision-making model which additionally considers minimizing the downtime of the system. By comparing the results of the two schemes, the decision-making scheme of NSGA-III is better than that of NSGA-II, which verifies that the three-objective decision-making model considering minimizing downtime can effectively reduce the downtime of the system.

Acknowledgement

This research is supported by Doctoral innovation fund of Xi'an University of Technology (No. 310-252072013), National Natural Science Foundation of China (Grant No. 52005404) and China Postdoctoral Science Foundation (No. 2020M673612XB).

References

[1] Zhang, Y.P., Yang, K.X., Shi, L. (2018). Economic optimization model for imperfect preventive maintenance under reliability constraints, *Computer Integrated Manufacturing Systems*, Vol. 24, No. 12, 3019-3026, doi: 10.13196/j.cims.2018.12.010.

[2] Laggoune, R., Chateauneuf, A., Aissani, D. (2010). Impact of few failure data on the opportunistic replacement policy for multi-component systems, *Reliability Engineering & System Safety*, Vol. 95 No. 2, 108-119, doi: 10.1016/j.ress.2009.08.007.

[3] Rashidnejad, M., Ebrahimnejad, S., Safari, J. (2018). A bi-objective model of preventive maintenance planning in distributed systems considering vehicle routing problem, *Computers & Industrial Engineering*, Vol. 120, 360-381, doi: 10.1016/j.cie.2018.05.001.

[4] Pandey, M., Zuo, M.J., Moghaddass, R., Tiwari, M.K. (2013). Selective maintenance for binary systems under imperfect repair, *Reliability Engineering & System Safety*, Vol. 113, 42-51, doi: 10.1016/j.ress.2012.12.009.

[5] Dao, C.D., Zuo, M.J., Pandey, M. (2014). Selective maintenance for multi-state series-parallel systems under economic dependence, *Reliability Engineering & System Safety*, Vol. 121, No. 1, 240-249, doi: 10.1016/j.ress.2013.09.003.

[6] Khatab, A., Aghezzaf, E.H., Diallo, C., Djelloul, I. (2017). Selective maintenance optimization for series-parallel systems alternating missions and scheduled breaks with stochastic durations, *International Journal of Production Research*, Vol. 55, No. 10, 3008-3024, doi: 10.1080/00207543.2017.1290295.

[7] Khatab, A. (2013). Hybrid hazard rate model for imperfect preventive maintenance of systems subject to random deterioration, *Journal of Intelligent Manufacturing*, Vol. 26, No. 3, 601-608, doi: 10.1007/s10845-013-0819-x.

[8] Chaabane, K., Khatab, A., Diallo, C., Aghezzaf, E.-H., Venkatadri, U. (2020). Integrated imperfect multimission selective maintenance and repairpersons assignment problem, *Reliability Engineering & System Safety*, Vol. 199, Article No. 106895, doi: 10.1016/j.ress.2020.106895.

[9] Cheng, G., Zhou, B., Li, L. (2019). Joint optimization of production quantity, quality control and predictive maintenance for production systems with multiple machines, *System Engineering – Theory & Practice*, Vol. 39, No. 8, 2152-2161, doi: 10.12011/1000-6788-2018-0104-10.

[10] Lu, Z., Zhao, C., Cui, W. (2017). Modeling and optimization of preventive maintenance scheduling for series production systems, *Journal of Harbin Engineering University*, Vol. 38, No. 2, 269-275, doi: 10.11990/jheu.201511036.

[11] Leng, Q., Wang, L., Zhao, S., Zheng, Y. (2016). Integration and optimization of production batch and maintenance plan for multi-machine serial system, *Computer Integrated Manufacturing System*, Vol. 22, No. 8, 1945-1952, doi: 10.13196/j.cims.2016.08.013.

[12] Li, Y.T., Li, J., Huang, Z.K. (2019). Opportunistic maintenance strategy for serial-parallel manufacturing system under multiple-resource constraint, *Journal of Lanzhou University of Technology*, Vol. 45, No. 5, 33-38.

[13] Bousdekis, A., Papageorgiou, N., Magoutas, B., Apostolou, D., Mentzas, G. (2017). A proactive event-driven decision model for joint equipment predictive maintenance and spare parts inventory optimization, *Procedia CIRP*, Vol. 59, 184-189, doi: 10.1016/j.procir.2016.09.015.

[14] Lopes Gerum, P.C., Altay, A., Baykal-Gürsoy, M. (2019). Data-driven predictive maintenance scheduling policies for railways, *Transportation Research Part C: Emerging Technologies*, Vol. 107, 137-154, doi: 10.1016/j.trc.2019.07.020.

[15] Xu, T.X., An, J., Zhang J.P., Li, Z.Q.(2018). Equipment condition maintenance decision-making model based on Markov approach, *Journal of Gun Launch & Control*, Vol. 39, No. 3, 90-94, doi: 10.19323/j.issn.1673-6524.2018.03.019.

[16] Pei, H., Hu, C.H., Si, X.S., Zhang, Z.X., Du, D.B. (2018). Equipment maintenance decision-making model based on remaining life prediction information under imperfect maintenance, *Acta Automatica Sinica*, Vol. 44, No. 4, 719-729, doi: 10.16383/j.aas.2017.c160534.

[17] Nguyen, K.-A., Do, P., Grall, A. (2015). Multi-level predictive maintenance for multi-component systems, *Reliability Engineering & System Safety*, Vol. 144, 83-94, doi: 10.1016/j.ress.2015.07.017.

[18] Dao, C.D., Zuo, M.J. (2017). Optimal selective maintenance for multi-state systems in variable loading conditions, *Reliability Engineering & System Safety*, Vol. 166, 171-180, doi: 10.1016/j.ress.2016.11.006.

[19] Srinivas, N., Deb, K. (1994). Muiltiobjective optimization using nondominated sorting in genetic algorithms, *Evolutionary Computation*, Vol. 2, No. 3, 221-248, doi: 10.1162/evco.1994.2.3.221.
[20] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, 182-197, doi: 10.1109/4235.996017.
[21] 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.
[22] Liu, Y.F., Zhang, Q.S. (2018). Solving multi-objective planning model for equipment manufacturing enterprises with dual uncertain demands using NSGA-II algorithm, *Advances in Production Engineering & Management*, Vol. 13, No. 2, 193-205, doi: 10.14743/apem2018.2.284.
[23] Zhong, Y., Li, J.M., Zhu, S.Z. (2017). Research on the multi-objective optimized scheduling of the flexible job-shop considering multi-resource allocation, *International Journal of Simulation Modelling*, Vol. 16, No. 3, 517-526, doi: 10.2507/IJSIMM16(3)CO13.

Calendar of events

- 35th Annual European Simulation and Modelling Conference, October 27-29, 2021, Rome, Italy.
- 32nd DAAAM International Symposium, October 28-29, 2021, Virtual conference.
- 6th International Conference on Manufacturing Technologies, January 22-24, 2022, Singapore.
- 11th International Conference on Operations Research and Enterprise Systems, February 3-5, 2022, Vienna, Austria.
- The 5th International Conference on Materials Engineering and Applications, February 12-14, 2022, Nha Trang, Vietnam.
- 9th International Engine Congress 2022, February 22-23, Baden-Baden, Germany, (hybrid event).
- 7th International Conference on Manufacturing, Material and Metallurgical Engineering, March 18-20, 2022, Osaka, Japan.
- The 10th IIAE International Conference on Industrial Application Engineering, March 26-30, 2022, Matsue, Japan (hybrid event).
- 50th SME North American Manufacturing Research Conference, June 27 to July 1, 2022, West Lafayette, USA.
- 16th International Conference on Micromachining Technology, October 17-18, 2022, Dubai, United Arab Emirates.

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 16 | Number 3 | September 2021 | pp 265-388

Contents

Published by CPE, University of Maribor

ISSN 1854-6250

9 771854 625008

apem-journal.org