

Yingsai CAO, Panfei WANG, Wenjie XV, Wenjie DONG

Joint optimization of quality-based multi-level maintenance and buffer stock within multi-specification and small-batch production

© Higher Education Press 2025

Abstract This study proposes a comprehensive framework for the joint optimization of maintenance actions and safety stock policies for multi-specification small-batch (MSSB) production. The production system considered consists of multiple machines arranged in a series-parallel configuration. Given the multi-stage nature of the MSSB, a piecewise Gamma process is developed to model the degradation of machines owing to varying product specifications. A quality-based maintenance model is proposed to guide the scheduling of maintenance actions based on the observed product defect rate. The maintenance policy is optimized at two levels: at the machine level, the optimal quality of the produced products is determined, and at the system level, a threshold quality value is established to facilitate the opportunistic maintenance of machines. The relationship between the buffer stock and machine capacity is explicitly modeled to ensure production efficiency. A simulation-based multi-objective algorithm is employed to identify the optimal decision variable levels for the proposed maintenance policy. The numerical results demonstrate that the proposed method effectively balances the conflicting objectives of minimizing the expected operational costs and maximizing production efficiency.

Keywords serial-parallel multi-stage production system, multi-specification and small-batch, multi-level maintenance, quality control, safety stock

1 Introduction

Maintenance significantly contributes to the overall operational costs as a critical component of operational management. According to Bevilacqua and Braglia (2000), maintenance costs account for up to 70% of total costs. Therefore, an increasing number of scholars are focusing on optimizing maintenance policies for complex production systems. As noted by Sun et al. (2023) and Hu et al. (2022), the former proposes a strategy that dynamically adjusts productivity to minimize maintenance costs in the least favorable scenarios, whereas the latter utilizes production downtime to schedule preventive maintenance (PM) replacements and applies approximate dynamic programming to optimize solutions.

For complex production systems with multiple machines, maintenance actions must often be scheduled at various levels. At the machine level, the threshold for PM is optimized. At the system level, maintenance actions for the machines involved can be further grouped to reduce costs. A comprehensive review of maintenance optimization at the machine level was conducted by de Jonge and Scarf (2020). When several machines are scheduled for maintenance together, intrinsic dependence can be utilized. As an effective strategy for reducing frequent downtime caused by maintenance, opportunistic maintenance (OM) aims to identify potential maintenance opportunities during the maintenance of a specific machine (Zhang and Yang, 2021). Similar to PM, optimal conditions exist for scheduling OM. Most existing research on this topic has focused on determining the optimal threshold for OM. For example, Zhou and Ning (2021) proposed an OM policy for multi-unit serial

Received Sep. 21, 2024; revised Jan. 24, 2025; accepted Feb. 17, 2025

Yingsai CAO (✉), Panfei WANG, Wenjie XV
School of Management, Jiangsu University, Zhenjiang 212013, China
E-mail: caoyingsai@foxmail.com

Wenjie DONG
College of Economics and Management, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China

This research was supported by the Natural Science Foundation of Jiangsu Province (No. BK20210739), Natural Science Foundation of Colleges and Universities in Jiangsu Province (No. 21KJB410002), Philosophy and Social Science Foundation of Colleges and Universities in Jiangsu Province (No. 2021SJA2049), Humanities and Social Science Fund of Ministry of Education of China (No. 24YJC630050).

production systems, explicitly modeling the influences of stochastic production waits on the OM threshold. Gan et al. (2022) jointly optimized threshold-based OM and production planning, with the objective of minimizing the total weighted expected completion time. The optimal threshold values for PM and OM were obtained simultaneously.

The research mentioned above provides valuable insights into enhancing the maintenance efficiency. However, nearly all these models assume stable production conditions. Therefore, these studies may better suit mass production rather than multi-specification small-batch (MSSB), which is characterized by variable production environments. Recent literature on maintenance optimization within an MSSB can be summarized as follows: Zhou et al. (2021) explicitly modeled the influence of uncertain demands on machine performance evaluation and maintenance scheduling. Zhu and Zhou (2022) addressed the variation in production plans in the MSSB and proposed a general model to estimate future failure rates. Maintenance actions are scheduled dynamically, allowing lifecycle information to be integrated. Zhu and Zhou (2023) simultaneously optimized spare part provision and maintenance planning, considering the variable working conditions arising from rapid changes in the products to be processed. An accelerated failure-time framework and a hierarchical-clustering-based OM policy were developed to ensure the operational efficiency of the MSSB. Zhou and Yu (2020) proposed a semi-dynamic maintenance policy for the production systems in MSSB. An effective OM policy was designed to respond to changes in production conditions.

Although the research presented can aid in analyzing variations in production speed and task loads within an MSSB, it heavily depends on the conditions assessed to determine whether maintenance actions should be scheduled. Frequent inspections may disrupt production processes, potentially further reducing the production efficiency.

To minimize production interruptions, several studies have focused on the quality of the products produced. This mechanism can be explained by the fact that, as machine performance declines, product quality decreases. Therefore, quality is often incorporated into maintenance optimization (Chen et al., 2023). Maintenance optimization may contain two objectives when product quality is considered: minimizing expected operational costs and maximizing production revenue (Shi et al., 2023).

A critical aspect of the joint optimization of maintenance and quality is examining the relationships between machine performance efficiency and product quality levels, which has attracted substantial attention. For instance, Cheng and Li (2020) explicitly modeled the relationship between machine deterioration and product quality in a multi-stage production system with series-parallel machines, allowing for the simultaneous

optimization of PM and quality control policies. Wang et al. (2024) investigated the interrelations among product quality, maintenance effort, and maintenance frequency within a servitization model, identifying optimal values for these factors to minimize total costs. Jain and Lad (2017) explored the relationship between product quality and tool degradation using real-time health-monitoring techniques. These findings support the dynamic scheduling of process quality control and maintenance. Azimpoor and Taghipour (2021) modeled the interactions between product quality and production planning, accounting for the delay time in the failure process. Boumallessa et al. (2023) employed a coupling function to illustrate the relationship between the machine degradation and product quality. Cheng et al. (2018) developed an effective model outlining the relationship between quality deterioration and machine reliability; however, this model is limited to production systems with a single product type. Wan et al. (2023) jointly optimized the economic production quantities, maintenance strategies, and quality control of a continuous flow manufacturing process to produce a single product. Taguchi's design of experiments (DOE) was used to assess the effects of machine degradation on product quality. Tasia (2022) framed the product quality within a multivariate production process and identified the influence of machine performance on this comprehensive quality level.

In addition to maintenance and quality control, inventory optimization plays a significant role in the MSSB. During machine maintenance, the production lines may need to be halted. Insufficient buffer stock may fail to meet the market demand, whereas excessive stock can create a cost burden. Thus, determining the safety buffer stock level is a key aspect of buffer optimization (Hadian et al., 2021). Mohammad Hadian et al. (2023) explored the interactions between maintenance and inventory control, simultaneously deriving optimal conditions for maintenance and safety stock levels. Liu et al. (2020) considered both imperfect PM and buffer inventory when optimizing the total operation costs. Zhu et al. (2024) focused on an onshore wind farm, jointly optimizing the maintenance and spare-part inventory policies. Polotski et al. (2022) developed a method to calculate the safety stock level of perishable finished goods.

The aforementioned studies provide valuable insights into maintenance policy design, product quality control, and buffer-stock optimization. However, nearly all of them were designed for large-lot production, overlooking the variety of product types to be processed. The MSSB production mode has gained popularity in manufacturing because of its benefits in enhancing flexibility and product quality. Customer demand often dictates frequent changes in product specifications and quantities within an MSSB. Therefore, the degradation rates of the production systems are not constant across different production tasks. Although typical models, such as the Proportional

Hazard Rate Model (Zhou and Yu, 2020) and the Accelerated Failure Time Model (Hu et al., 2020), can support decision-making in reliability evaluation under varying operational conditions, the influence of MSSB on multi-level maintenance scheduling and quality optimization remains unaddressed.

To address these challenges, this study proposes a quality-based multi-level maintenance approach. Based on the observed quality information, the maintenance decisions are optimized at both the machine and overall production system levels. The impact of product differences on maintenance and buffer stock policies is also examined.

The potential contributions of this study can be summarized as follows. 1) The influences of different product specifications on machine reliability are explicitly modeled through a multi-stage gamma process. 2) A quality-based maintenance model is proposed to specify how multilevel maintenance actions can be scheduled based on the inspected quality information. 3) A simulation-based multi-objective Non-dominated Sorting Genetic Algorithm II (NSGA-II) is designed to determine the optimal level of decision variables related to the proposed maintenance and buffer stock policy.

The rest of this paper is organized as follows. The structural characteristics of the production systems considered for MSSB, quality-based multilevel maintenance, and buffer stock policy are presented in Section 2. Section 3 constructs the joint optimization model of the proposed maintenance and buffer stock policies. Section 4 presents the design of a simulation-based multiobjective NSGA-II algorithm to obtain the optimal values of the decision variables. Section 5 demonstrates the effectiveness and practicality of the proposed method using an illustrative example. Section 6 summarizes the conclusions and discusses future research interests.

2 Maintenance modeling and buffer stock policy

2.1 Production system for MSSB

This study considers the MSSB within order-driven production (pull system). A multi-stage serial-parallel production system is introduced to achieve MSSB. We assume that there are N_s orders, each corresponding to a specific product type. The production rate is set as P_s . Let the production cycle of the s th ($1 \leq s \leq N_s$) order be l_s which refers to the duration from the start of the s th order to the start of the $(s+1)$ th order. There are multiple stages in l_s . The order process cycle is assumed to follow a uniform distribution, i.e., $l_s \sim U(\text{Min}_{l_s}, \text{Max}_{l_s})$. Multiple machines are employed to complete the operations in every stage. Diagram of the considered production system is shown in Fig. 1. There are N stages in the system. k_m represents the total number of machines in the m th ($1 \leq m \leq N$) stage. M_{smj} represents the j th ($1 \leq j \leq k_m$) machine in the m th stage of the s th order.

The sum of the processing rates of all machines in each stage equals to the total production rate to satisfy the line balancing condition, i.e., $\sum_{j=1}^{k_m} p_{smj} = P_s$ where p_{smj} is the processing capacity of M_{smj} . Each machine deteriorates randomly during production. Degradation level of machine M_{smj} at time t is represented as $X_{smj}(t)$. This indicates that the production order directly influences how the machines degrade. The underlying mechanism can be explained as products in different orders may not be the same within MSSB. Furthermore, the manufacturing requirements for different products may vary, and the effects of the products being processed on machines can differ based on product types in each production order. The Gamma process is utilized to capture the stochastic

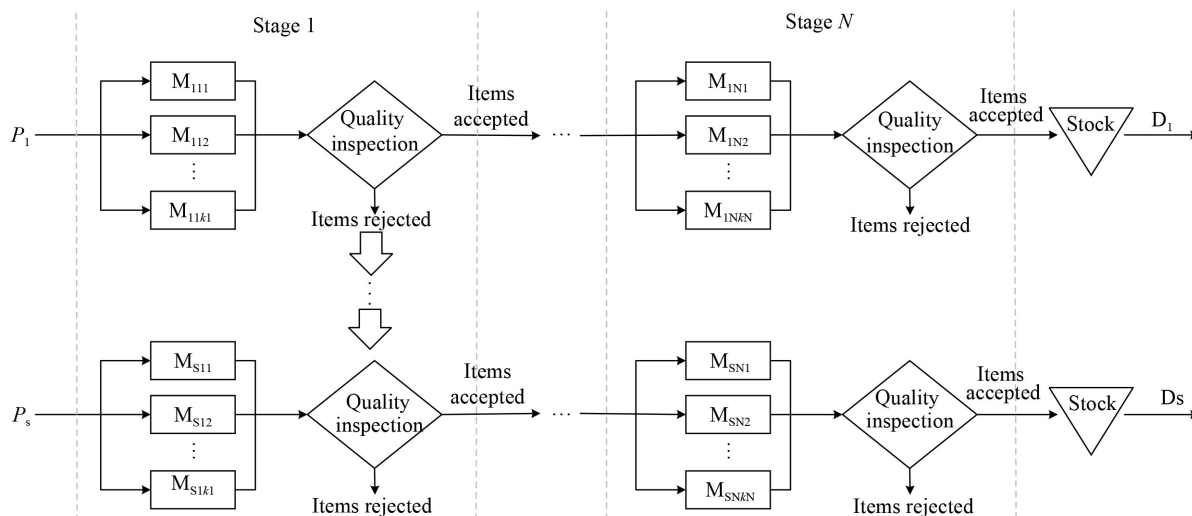


Fig. 1 Structure of the multi-stage series-parallel system for MSSB.

characteristics due to its advantages in modeling stationary and independent degradations (Cao et al., 2024). Within Gamma process, the degradation increment in the time interval $[t_1, t_2]$ follows a Gamma distribution with shape parameter $\alpha_{smj}(t_2 - t_1)$ and scale parameter β_{smj} . The corresponding probability density can be expressed as

$$f_{smj}(x, t_2 - t_1) = \frac{\beta_{smj}^{\alpha_{smj}(t_2 - t_1)} x^{\alpha_{smj}(t_2 - t_1) - 1}}{\Gamma[\alpha_{smj}(t_2 - t_1)]} e^{-\beta_{smj}x}, x \geq 0, \quad (1)$$

where $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du, a > 0$ is the Gamma function. The degradation rate can be obtained as α_{smj}/β_{smj} . Failure occurs when the degradation level $X_{smj}(t)$ at time t exceeds the failure threshold L_{smj} .

Let τ_{smj} be the first failure time of machine M_{smj} , i.e., $\tau_{smj} = \inf\{t | M_{smj}(t) \geq L_{smj}\}$. The failure probability of machine M_{smj} at time t can be formulated as

$$\begin{aligned} F_{\tau_{smj}}(t) &= Pr(\tau_{smj} < t) = Pr(X_{smj}(t) \geq L_{smj}) \\ &= \int_{L_{smj}}^\infty f_{smj}(x, t) dx = \int_{L_{smj}}^\infty \frac{\beta_{smj}^{\alpha_{smj}t} x^{\alpha_{smj}t - 1}}{\Gamma(\alpha_{smj}t)} e^{-\beta_{smj}x} dx, \\ &\xrightarrow{\text{set } u = \beta_{smj}x} \int_{L_{smj}\beta_{smj}}^\infty \frac{u^{\alpha_{smj}t - 1}}{\Gamma[\alpha_{smj}t]} e^{-u} du \\ &= \frac{\Gamma(\alpha_{smj}t, L_{smj}\beta_{smj})}{\Gamma(\alpha_{smj}t)}, \end{aligned} \quad (2)$$

where $\Gamma(a, b)$ is the incomplete Gamma function, defined as $\Gamma(a, b) = \int_b^\infty u^{a-1} e^{-u} du, a > 0, b > 0$.

Then, reliability of machine M_{smj} at time t can be obtained as $R_{smj}(t) = 1 - F_{\tau_{smj}}(t)$.

Customer demand determines that product specifications and urgency may vary with the production tasks. Therefore, the machines involved may degrade through various mechanisms. The shape parameter of the Gamma process is employed to explicitly model the influences of different product specifications. More specifically, the shape parameter of the j th machine in the m th stage within the s th order at time t can be formulated as:

$$\alpha'_{smj}(t) = \alpha_{0mj}(t) \cdot \exp(b_{1mj}d_{smj} + b_{2mj}q_{smj}), \quad (3)$$

where $\alpha_{0mj}(t)$ represents the shape parameter level at the initial stage where no product differences exist. d_{smj} and q_{smj} represent the ratio of the manufacturing process and processing intensity requirements of the processing condition in the s th order to the initial stage (when $s = 0$), respectively. b_{1mj} and b_{2mj} are the influencing parameters. $\exp(b_{1mj}d_{smj} + b_{2mj}q_{smj})$ can be viewed as the comprehensive influence brought by changes of the s th order in the aspects of manufacturing process and processing intensity.

2.2 Quality-based multi-level maintenance

To maintain the operational efficiency of the production

system, a quality-based multi-level maintenance approach is designed in this section. First, maintenance actions at the machine level include PM, corrective maintenance, and overhaul. Maintenance at the system level corresponds to OM, which groups maintenance actions for several machines, resulting in lower costs compared to maintaining them separately. Second, observed quality levels of the output products are used to evaluate the performance levels of the involved machines. When the inspected quality level exceeds a threshold, PM is scheduled at the machine level. Additionally, during production operations, machines may fail unexpectedly. Corrective maintenance actions are planned for such situations. Both corrective maintenance and overhaul restore machines to their original condition. The durations of preventive, opportunistic, and corrective maintenance activities are considered negligible. The paper adopts the No-Return (NR) policy (Chiu et al., 2012), indicating that new production orders will be initiated once the existing safety stock is depleted.

2.2.1 Quality-based maintenance in machine level

Within the MSSB, quality plays a significant role in market competition. To satisfy customer demand, thorough quality inspections are conducted at each stage to prevent defective products from moving to the next stage. The defect rate of products from each machine can be monitored in real time. Therefore, it may not be necessary to stop machines from measuring their actual degradation levels. Take the machine M_{smj} For example, according to Bouslah et al. (2016), the relationship between the underlying degradation level and observed defect rate can be expressed as

$$X_{smj}(t) = \left[\frac{\ln \eta_{smj} - \ln(\eta_{smj} - \tilde{p}_{smj}(t) + \tilde{p}_{0mj})}{\lambda_{smj}} \right]^{\frac{1}{\gamma_{smj}}}, \quad (4)$$

where $\tilde{p}_{smj}(t)$ refers to the real-time product defect rate from machine M_{smj} at time t , \tilde{p}_{0mj} refers to the initial defect rate of the output products when the machine is brand-new. η_{smj} is the boundary value of the product quality. λ_{smj} and γ_{smj} are two positive variables. Based on the historical data, the maximum likelihood method can be employed to evaluate the above parameters.

Once the machine degradation level is assessed from Eq. (4), maintenance actions can be scheduled based on the observed conditions. More specifically, when the real-time defect rate of production from a specific machine exceeds the quality threshold QT , the PM is scheduled.

In addition to PM, overhauls are scheduled to reduce the failure risks of machines as much as possible. The performance level of each machine is inspected after each order. This inspection is thorough and incurs a specific expected cost. Based on the inspection results, overhauls

can be scheduled when necessary. The degradation level of the machine resets to 0 after overhauls, while the degradation rate remains unchanged. Also, take the machine M_{smj} for example, on the condition that the inspected degradation level is x_{smj} at the end of the s th order, reliability level in the $(s+1)$ th order can be predicted as

$$\begin{aligned} R_{smj}^{(s+1)}(l_s|x_{smj}) &= \Pr[X_{smj}(t_{s+1}) < L_{smj}|X_{smj}(t_s) = x_{smj}] \\ &= \Pr[X_{smj}(t_s) + X_{smj}(l_s) < L_{smj}|X_{smj}(t_s) = x_{smj}] \\ &= 1 - \Pr[X_{smj}(l_s) \geq L_{smj} - x_{smj}] \\ &= 1 - \frac{\Gamma(\alpha_{smj} \cdot \exp(b_{1mj}d_{smj} + b_{2mj}q_{smj}) \cdot l_s, (L_{smj} - x_{smj})\beta_{smj})}{\Gamma(\alpha_{smj} \cdot \exp(b_{1mj}d_{smj} + b_{2mj}q_{smj}) \cdot l_s)}. \end{aligned} \quad (5)$$

To optimize the use of available maintenance resources, the structural importance and machine capacity rate are considered alongside the predicted reliability level. Specifically, machines with higher structural importance and production capacity should be prioritized for scheduling overhauls. The overhaul threshold for a machine M_{smj} can be formulated as:

$$\begin{aligned} \psi_{smj} &= W \cdot IB_{mj} \cdot \frac{P_{smj}}{\max\{p_{smj} | 1 \leq j \leq k_m\}} \\ &= W \cdot IB_{mj} \cdot CR_{smj} \end{aligned} \quad (6)$$

where W is the decision variable. IB_{mj} represents the structural importance of machine M_{smj} . This can be calculated according to (Kuo and Zhu, 2012). $CR_{smj} = p_{smj}/\max\{p_{smj} | 1 \leq j \leq k_m\}$ is the ratio of the production capacity of machine M_{smj} to the maximum production capacity in the m th stage. If the predicted reliability $R_{smj}^{(s+1)}(l_s|x_{smj})$ is less than the threshold ψ_{smj} , the machine is scheduled to overhaul.

2.2.2 Opportunistic maintenance in system level

Owing to economic dependence, maintaining several machines may be less costly than repairing them separately. This is attributed to setup activities, such as spare parts provision and maintenance worker scheduling. In addition, maintaining a specific machine may create opportunities for other machines to undergo PM. In this section, production efficiency is considered to evaluate the priority of OM. The OM threshold of machine M_{smj} can be expressed as:

$$\omega_{smj} = QT \cdot (1 - H \cdot CR_{smj}), \quad (7)$$

where H is a decision variable to be optimized.

Similar to PM, OM cannot restore the machine to the brand-new state. The effects of both PM and OM

decreased with the number of already conducted PMs and OMs.

An accelerated gamma process was introduced to comprehensively describe the degradation process within multilevel maintenance. Inherent stochastic monotonicity can be employed to model the increasing failure rate mechanism. Let $\tau_{smj}^{(i)}$ represent the first failure time of machine M_{smj} after the $(i-1)$ th preventive OM. Subsequently, $\{\tau_{smj}^{(i)}, i = 1, 2, 3, \dots\}$ forms a randomly decreasing geometric process. The cumulative distribution function of $\tau_{smj}^{(i)}$ can be formulated as

$$F_{\tau_{smj}^{(i)}}(t) = \frac{\Gamma(a_{smj}^{i-1} \cdot \alpha_{smj} \cdot \exp(b_{1mj}d_{smj} + b_{2mj}q_{smj}) \cdot t, L_{smj}\beta_{smj})}{\Gamma(a_{smj}^{i-1} \cdot \alpha_{smj} \cdot \exp(b_{1mj}d_{smj} + b_{2mj}q_{smj}) \cdot t)}, a_{smj} > 1. \quad (8)$$

The shape and scale parameters of the accelerated Gamma process after the $(i-1)$ th maintenance are $a_{smj}^{i-1} \cdot \alpha_{smj} \cdot \exp(b_{1mj}d_{smj} + b_{2mj}q_{smj})$ and β_{smj} . Because $a_{smj} > 1$, the machine degradation rate increases with the maintenance time i .

2.3 Buffer stock policy

As an effective method for ensuring production supply, buffer stock policy has been widely employed to rapidly satisfy customer demand and improve productivity (Hadian et al., 2021). The optimization result of buffer stock is the total number of stocked products, which can prevent product shortages during planned maintenance actions (Lopes, 2018). Within MSSB, buffer stock is scheduled based on the types of involved orders due to variations in customer demands. For all machines, the initial stock level is sufficient (referred to as safety stock). If the duration of an overhaul is within the time needed to consume all available stock, product shortages will not occur. The production rate and processing intensity can improve with an increase in safety stock. Due to differences in product specifications, machine production capacity may vary at different stages. Therefore, the safety stock level differs among machines. On the condition that the total safety stock of the production system for the s th order is represented by SS (decision variable), the safety stock of machine M_{smj} can be expressed as:

$$S_{smj} = SS \cdot h_{smj}, \quad (9)$$

where $h_{smj} = p_{smj}/P_s$ represents the ratio of the capacity of machine M_{smj} to the total capacity for the whole processing stage.

Taking machine M_{smj} as an example, if it is under overhaul, customer demand is met by the buffer stock. Then, the stock level decreases at the rate of p_{smj} . If the repair time exceeds the time required for the stock level to drop to zero, a product shortage occurs.

3 Joint optimization of quality-based multi-level maintenance and buffer stock

To obtain the optimal values of decision variables in quality-based maintenance and buffer stock policy, optimization objectives are specified as minimizing the production cost rate and maximizing the effective time rate. Cost rate is a common factor measuring the expected cost level of operation and has been widely employed for optimizing maintenance and production. Effective time rate is considered in this paper due to its significant contributions to production efficiency, which plays an important role in MSSB. It is defined as the ratio of the time required to produce qualified products to the total production time.

3.1 Expected cost rate calculation

The expected total cost for the s th order is denoted as TC_s which can be formulated as

$$TC_s = N_s \cdot C_{set} + C_{in} + DC_s + SC_s + IC_s + OHC_s + PMC_s + OMC_s + CMC_s, \quad (10)$$

where C_{set} represents the setup cost before each order. C_{in} is the inspection cost. DC_s refers to the expected cost caused by defective products. SC_s is the expected cost induced by stockout. IC_s is the expected inventory cost. OHC_s is the overhaul cost. PMC_s is the expected cost for PM actions. OMC_s is the expected cost in OM. CMC_s represents the expected cost by corrective maintenance.

3.1.1 Expected defective product loss cost for the s th order

Defective products may arise at each processing stage. Thus, the total number of defective products in the entire system is the sum of the defective products at each stage.

Take the s th order for example, the input rate of raw materials in the first stage is set as P_s , and its output rate is denoted by $p_{s1}(t)$. The input material of the second stage is the output product of the first stage. Then, $p_{sm}(t)$ can then be viewed as the output rate of the m th stage. As shown in Fig. 2, the output rate of the final stage is represented by $p_{sN}(t)$.

The output rate of the first stage is the sum of the output rates of all machines in this stage. The expression can be formulated as

$$\begin{aligned} p_1(t) &= P_s - [p_{s11} \cdot \tilde{p}_{s11}(t) + p_{s12} \cdot \tilde{p}_{s12}(t) + \dots + p_{s1k_1} \cdot \tilde{p}_{s1k_1}(t)] \\ &= P_s \cdot \left\{ 1 - \left[\frac{p_{s11}}{P_s} \cdot \tilde{p}_{s11}(t) + \frac{p_{s12}}{P_s} \cdot \tilde{p}_{s12}(t) + \dots + \frac{p_{s1k_1}}{P_s} \cdot \tilde{p}_{s1k_1}(t) \right] \right\} \\ &= P_s \cdot \{ 1 - [h_{s11} \cdot \tilde{p}_{s11}(t) + h_{s12} \cdot \tilde{p}_{s12}(t) + \dots + h_{s1k_1} \cdot \tilde{p}_{s1k_1}(t)] \} \\ &= P_s \cdot [1 - \mathbf{h}_{s1} \cdot \tilde{\mathbf{P}}_{s1}(t)^T], \end{aligned} \quad (11)$$

where $\mathbf{h}_{s1} = (h_{s11}, h_{s12}, \dots, h_{s1k_1})$. $h_{smj} = p_{smj}/P_s$ is the ratio of the capacity of machine M_{smj} in the s th order to the total capacity of production stage. $\tilde{\mathbf{P}}_{s1}(t) = (\tilde{p}_{s11}(t), \tilde{p}_{s12}(t), \dots, \tilde{p}_{s1k_1}(t))$ is the quality degradation vector of each machine in the first stage of the s th order.

Then, the output rate of the second stage can be expressed as

$$\begin{aligned} p_2(t) &= p_1(t) \cdot [1 - \mathbf{h}_{s2} \cdot \tilde{\mathbf{P}}_{s2}(t)^T] \\ &= P_s \cdot [1 - \mathbf{h}_{s1} \cdot \tilde{\mathbf{P}}_{s1}(t)^T] \cdot [1 - \mathbf{h}_{s2} \cdot \tilde{\mathbf{P}}_{s2}(t)^T]. \end{aligned} \quad (12)$$

Similarly, the output rate of the stage N can be derived as $p_{sN}(t) = P_s \cdot \prod_{i=1}^N [1 - \mathbf{h}_{si} \cdot \tilde{\mathbf{P}}_{si}(t)^T]$. The defective output rate of the whole production system at time point t can be calculated as

$$\tilde{P}_s(t) = P_s - p_{sN}(t) = P_s \cdot \left\{ 1 - \prod_{i=1}^N [1 - \mathbf{h}_{si} \cdot \tilde{\mathbf{P}}_{si}(t)^T] \right\}. \quad (13)$$

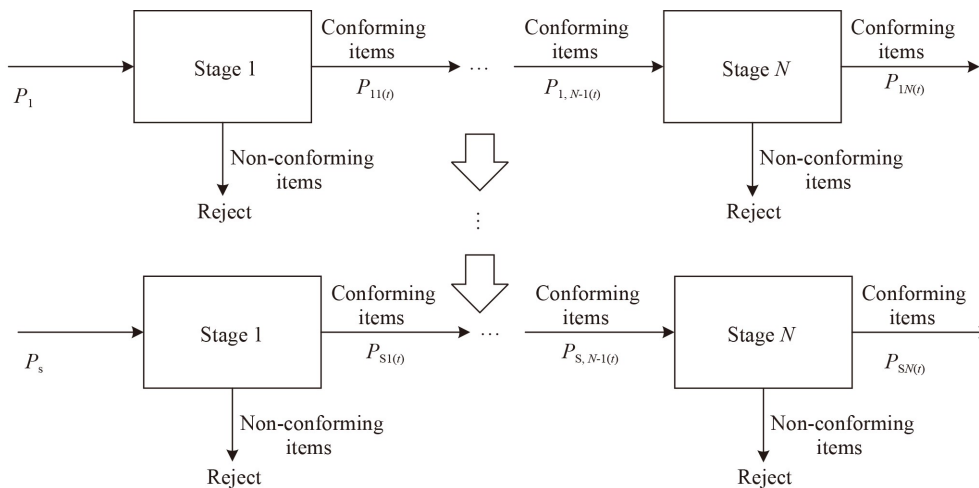


Fig. 2 Diagram of material flow.

Let the start time of the s th production order be t_s , then the expected defective product loss can be formulated as

$$DC_s = c_d \cdot \int_{t_s}^{t_s+l_s} [\tilde{P}_s(t)] dt = c_d \cdot P_s \cdot \int_{t_s}^{t_s+l_s} \left\{ 1 - \prod_{i=1}^N [1 - \mathbf{h}_{si} \cdot \tilde{P}_{si}(t)^T] \right\} dt. \tag{14}$$

3.1.2 Expected inventory cost for the s th order

Let $I_{smj}(t)$ be the inventory level of machine M_{smj} in the s th production order. The initial inventory of the production system is assumed to be the system-level safety stock S_s . The initial inventory level of M_{smj} is the machine-level safety stock S_{smj} . Within a specified order, the inventory holding cost can be subdivided into three components: cost during the normal production stage, cost during maintenance-induced downtime, and cost incurred from the end of maintenance to the restoration of normal production.

In the normal stage, the production speed remained stable at p_{smj} . The demand rate at this stage is equal to the output rate; thus, the inventory level satisfies $I_{smj}(t) = S_{smj}$. The expected inventory-holding cost for machine M_{smj} can be calculated as $c_h \cdot \int_{t_s}^{t_s+l_s} [I_{smj}(t)] dt$. Then, the inventory holding cost of the entire production system can be formulated as $\sum_{(m,j) \in G_s} \left\{ c_h \cdot \int_{t_s}^{t_s+l_s} [I_{smj}(t)] dt \right\}$ where G_s is the set of involved machines in the whole production system in the s th order.

During the downtime caused by maintenance, machines must stop production. The inventory level decreases from the initial level S_{smj} . The time-dependent inventory of machine M_{smj} can be expressed as $I_{smj}(t) = S_{smj} - p_{smj} \cdot t$ where p_{smj} represents the rate of decrease. The expected inventory holding cost at this stage is $c_h \cdot \int_0^{t_{IC}^1} (S_{smj} - p_{smj} \cdot t) dt$ where t_{IC}^1 represents the duration of the stage. If the overhaul time t_{smj}^{ohm} is greater than the consumption time of safety stock S_{smj}/p_{smj} , then $t_{IC}^1 = S_{smj}/p_{smj}$. Stockout occurs under the aforementioned conditions. The expected stockout cost is presented in Section 3.1.3. If the overhaul time t_{smj}^{ohm} is not larger than the consumption time of the safety stock S_{smj}/p_{smj} , then $t_{IC}^1 = t_{smj}^{ohm}$. To summarize, the expected inventory-holding cost of the entire production system in this stage can be expressed as $\sum_{(m,j) \in G_s} \left\{ c_h \cdot \int_0^{t_{IC}^1} (S_{smj} - p_{smj} \cdot t) dt \right\}$.

In the stage from the end of maintenance to the return of normal production, the machine resumes operations. The production rate gradually increases as the inventory level rises. Once the inventory level reaches the safety stock, the production rate returns to the normal stage. On the condition that the inventory change rate is p'_{smj} , inventory level of machine M_{smj} at time t can be expressed as $I_{smj}(t) = p'_{smj} \cdot t$. Then, the expected inventory

holding cost can be formulated as $c_h \cdot \int_0^{t_{IC}^2} (p'_{smj} \cdot t) dt$ where t_{IC}^2 is the duration of this stage and it can be calculated as $t_{IC}^2 = S_{smj}/p'_{smj}$. The expected inventory holding cost of the entire production system during this stage can be calculated as $\sum_{(m,j) \in G_s} \left\{ c_h \cdot \int_0^{S_{smj}/p'_{smj}} (p'_{smj} \cdot t) dt \right\}$.

The total inventory holding the expected cost for the s th order can be formulated as

$$IC_s = \sum_{(m,j) \in G_s} \left\{ c_h \cdot \int_{t_s}^{t_s+l_s} [I_{smj}(t)] dt \right\} + \sum_{(m,j) \in G_s} \left\{ c_h \cdot \int_0^{t_{IC}^1} (S_{smj} - p_{smj} \cdot t) dt \right\} + \sum_{(m,j) \in G_s} \left\{ c_h \cdot \int_0^{S_{smj}/p'_{smj}} (p'_{smj} \cdot t) dt \right\}. \tag{15}$$

3.1.3 Expected shortage cost for the s th order

When the overhaul time t_{smj}^{ohm} is greater than the consumption time of the safety stock S_{smj}/p_{smj} , a product shortage occurs. The corresponding shortage loss can be calculated as $c_s \cdot \int_0^{t_{smj}^{ohm} - S_{smj}/p_{smj}} (p_{smj} \cdot t) dt \cdot \chi\{t_{smj}^{ohm} > S_{smj}/p_{smj}\}$, where $\chi\{Z\}$ is the indicator function. If event Z occurs, the value of $\chi\{Z\}$ is 1; otherwise, it is zero. Furthermore, the expected shortage cost for the entire production system can be expressed as

$$SC_s = c_s \cdot \sum_{(m,j) \in G_s^{OH}} \left\{ \int_0^{t_{smj}^{ohm} - S_{smj}/p_{smj}} (p_{smj} \cdot t) dt \cdot \chi\{t_{smj}^{ohm} > S_{smj}/p_{smj}\} \right\}, \tag{16}$$

where G_s^{OH} is the set of subscripts of machines undergoing overhaul at the end of the s th order, i.e. $G_s^{OH} = \{(m, j) | R_{smj}^{(s+1)}(l_s | x_{smj}) < \psi_{smj}, t_s < t < t_s + l_s\}$.

3.1.4 Expected PM cost for the s th order

Based on the 100% quality inspection, the time-dependent defective rate of each involved machine can be obtained. A machine will be scheduled for PM if the defective rate reaches the quality control threshold QT .

Let the set G_s^{PM} be the set of machines of which defective rates exceed QT in the s th order, i.e. $G_s^{PM} = \{(m, j) | \tilde{p}_{smj}(t) \geq QT, t_s < t < t_s + l_s\}$. The expected PM cost for the entire production system in the s th order can then be calculated as $PMC_s = \sum_{(m,j) \in G_s^{PM}} \{C_{PM}^{mj}\}$.

3.1.5 Expected OM cost for the s th order

If the inspected defective rate of machine M_{smj} reaches its OM threshold ω_{smj} when other machines undergo PM, it is scheduled to be OM. Let G_s^{OM} be the set of machines which require OM in the i th PM of the s th order, i.e.

$G_s^{OM} = \{(m, j) | QT > \tilde{p}_{s,mj}(t) \geq \omega_{smj}, t_s < t < t_s + l_s\}$ where $\tilde{p}_{s,mj}(t)$ represents the defective rate of machine M_{smj} at the i th PM of the s th order. Then, the expect OM cost of the whole production system in the s th order can be calculated as $OMC_s = \sum_{i=1}^{Num(G_s^{PM})} \sum_{(m,j) \in G_s^{OM}} \{C_{OM}^{mj}\}$, where $Num(G)$ represents the number of elements in the set G .

3.1.6 Expected corrective maintenance cost for the s th order

We define G_s^{CM} as the set of machines which experienced random failures during the production period of the s th order, i.e., $G_s^{CM} = \{(m, j) | X_{smj}(t) \geq L_{smj}, t_s < t < t_s + l_s\}$. The expected corrective maintenance cost for the entire production system of the s th order can be calculated as $CMC_s = \sum_{(m,j) \in G_s^{CM}} \{C_{CM}^{mj}\}$.

3.1.7 Expected overhaul cost for the s th order

At the end of the s th order, each machine will be inspected to specify its current degradation level and to

predict the future reliability level. If the predicted reliability $R_{smj}^{(s+1)}(l_s | x_{smj})$ is below the overhaul threshold ψ_{smj} , the machine is scheduled to be an overhaul. The set of machines that require overhauling at the end of the s th order is denoted as G_s^{OH} . Thus, the expected overhaul cost for the entire production system can be expressed as $OHC_s = \sum_{(m,j) \in G_s^{OH}} \{C_{OH}^{mj}\}$.

Above all, the expected cost rate can be formulated as

$$C(W, QT, H, SS) = \frac{1}{\sum_{s=1}^{N_s} l_s} \cdot \sum_{s=1}^{N_s} TC_s. \quad (17)$$

3.2 Effective time utilization function

This section employs effective time rate to quantify maintenance efficiency. It can be defined as the ratio of the time required to produce qualified products to the total time required for production. First, the effective time ET_s refers to the time needed to produce qualified products during the s th order. It can be formulated as

$$ET_s = \frac{\left\{ \left[l_s \cdot P_s - \int_{t_s}^{t_s+l_s} [\tilde{P}_s(t)] dt \right] \cdot l_s \cdot N_{(m,j)} - \sum_{(m,j) \in G_s^{OH}} \left[\int_{t_s}^{t_s+l_s} (t_{smj}^{ohm}) dt \right] \right\}}{N_{(m,j)}}. \quad (18)$$

where $N_{(m,j)}$ represents the total number of machines in the production system. $\tilde{P}_s(t)$ is the defective rate of the entire production system at time t .

Therefore, the effective time utilization rate of the entire production system during the s th order production period can be calculated as $rET_s = ET_s / l_s$. Then, the expected effective time utilization rate can be expressed as

$$RET = \sum_{s=1}^{N_s} \left(rET_s \cdot \frac{l_s}{\sum_{s=1}^{N_s} l_s} \right). \quad (19)$$

To summarize, the joint optimization model can be constructed as

$$\begin{aligned} & \min C(W, QT, H, SS), \\ & \max RET(W, QT, H, SS), \\ \text{s.t.} & \begin{cases} 0 < W < \max\{(IB_{mj} \cdot CR_{smj})^{-1}\}, \\ 0 < QT < 1, \\ 0 < H < \max\{(CR_{smj})^{-1}\}, \\ SS \in \mathbb{Z}^+, t_s < t < t_s + l_s. \end{cases} \end{aligned} \quad (20)$$

4 Model solution

As shown in Eq. (20), the objective functions are nonlinear and include several stochastic variables, such as machine degradation increments, total types of orders, and the total number of maintenance actions conducted. Analytical expressions for these two objective functions may be difficult to obtain. In this section, we propose a simulation-based optimization algorithm to determine the optimal values of the decision variables.

4.1 Simulation of quality-based multi-level maintenance

In this section, the Monte Carlo technique is employed to simulate machine degradation, durations of different orders, and characteristics related to maintenance actions. The diagram of the entire simulation is shown in Fig. 3. The detailed procedure is as follows.

Step 1. Parameter initialization. Production order parameters: $s, N_s, l_s, d_{smj}, q_{smj}$. Production system parameters: $P_s, D_s, N, m, k_m, p_{smj}, \alpha_{smj}, \beta_{smj}, CR_{smj}, X_{smj} = 0, L_{smj}, IB_{mj}$. Quality-related parameters: $\tilde{p}_{smj}(t), \tilde{p}_{0mj}, \eta_{smj}, \lambda_{smj}, \gamma_{smj}$. Maintenance-related parameters: t_{smj}^{ohm}, a_{smj} . Cost-related parameters: $C_{set}, C_{in}, c_d, c_h, c_s, C_{PM}^{mj}, C_{OM}^{mj}, C_{CM}^{mj}, C_{OH}^{mj}$. Decision parameters: W, QT, H, SS . Simulation parameters: $t, s = 0, T_{total} = 0, C_{total} = 0,$

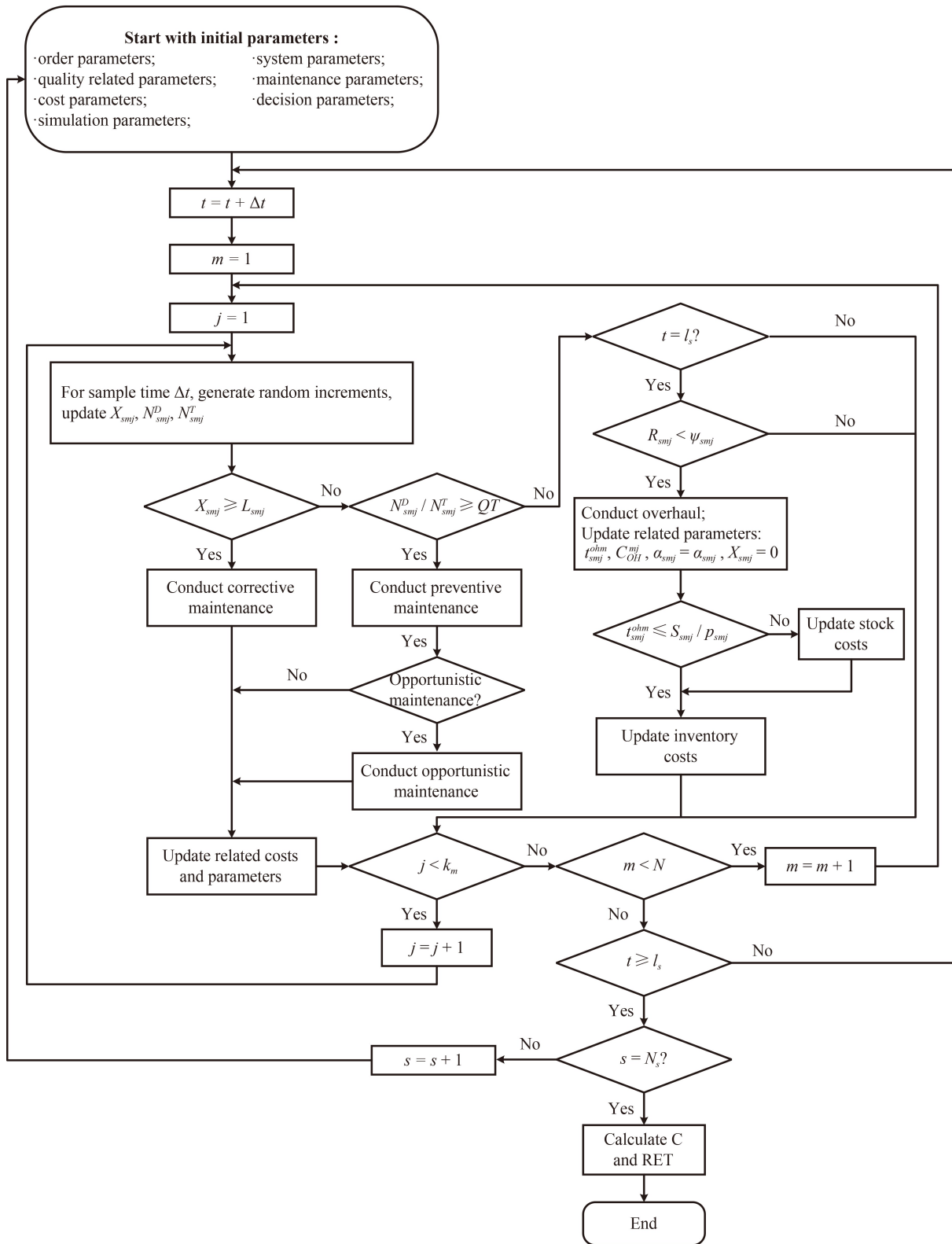


Fig. 3 flowchart of Monte Carlo simulation.

$ET_s = 0$.

Step 2. Simulate the degradation process of each piece of machine. Let $t = 0$, $s = 1$, $N_{smj}^D = 0$ (The number of

defective products produced by machine M_{smj} in the s th order), $N_{smj}^T = 0$ (The total number of products produced by machine M_{smj} in the s th order), $I_{smj} = S_{smj}$ (The inven-

tory level of machine M_{smj} in the s th order), $DC_s = 0$, $SC_s = 0$, $IC_s = 0$, $OHC_s = 0$, $PMC_s = 0$, $OMC_s = 0$, $CMC_s = 0$, $T_s^{ohm} = 0$ (The total overhaul time in the s th order), $s = s + 1$.

For the time increment Δt , $t = t + \Delta t$, machine M_{smj} will have a random deterioration increment $\Delta X_{smj}(\Delta t) = random('gamma', \alpha_{smj}\Delta t, \beta_{smj})$, then the cumulative amount of machine deterioration is $X_{smj} = X_{smj} + \Delta X_{smj}(\Delta t)$.

Step 3. Calculate defective product losses and inventory expected costs. The number of defective products produced by machine M_{s1j} during stage 1 of order s : $N_{s1j}^D = N_{s1j}^D + P_s \cdot h_{s1j} \cdot \tilde{p}_{s1j}(t) \cdot \Delta t$, where h_{s1j} represents the ratio of the capacity of machine M_{s1j} to the total capacity of all machines in stage 1, and $\tilde{p}_{s1j}(t)$ represents the defective product output rate of machine M_{s1j} at time t . Next, we calculate the number of defective products produced by machine M_{s2j} in the second stage: $N_{s2j}^D = N_{s2j}^D + p_1(t) \cdot h_{s2j} \cdot \tilde{p}_{s2j}(t) \cdot \Delta t$, where $p_1(t) = P_s \cdot [1 - h_{s1} \cdot \tilde{P}_{s1}(t)^T]$ is the overall output rate of all machines in stage 1, as detailed in Eq. (11). Similarly, the number of defective products produced by machine M_{smj} during stage m of order s is $N_{smj}^D = N_{smj}^D + p_{m-1}(t) \cdot h_{smj} \cdot \tilde{p}_{smj}(t) \cdot \Delta t$, ($1 \leq m \leq N$), where $p_{m-1}(t) = P_s \cdot \prod_{i=1}^{m-1} [1 - h_{si} \cdot \tilde{P}_{si}(t)^T]$ is the overall output rate of all machines in stage $(m-1)$. Calculate the defective product loss $DC_s = DC_s + \sum_{m=1}^N \sum_{j=1}^{k_m} c_d \cdot N_{smj}^D$. Similarly, the total number of products manufactured by machine M_{smj} at time t is $N_{smj}^T = N_{smj}^T + P_s \cdot \prod_{i=1}^{m-1} [1 - h_{si} \cdot \tilde{P}_{si}(t)^T] \cdot h_{smj} \cdot \Delta t$, ($1 \leq m \leq N$).

Update the inventory level $I_{smj} = I_{smj}$; if an overhaul occurs, I_{smj} will change, as detailed in step 7. Therefore, the expected inventory cost: $IC_s = IC_s + \sum_{m=1}^N \sum_{j=1}^{k_m} c_h \cdot I_{smj} \cdot \Delta t$.

Step 4. Conduct corrective maintenance and calculate related expected costs. Compare the deterioration amount of the machine with the failure threshold. If $X_{smj} \geq L_{smj}$, machine M_{smj} will fail, and then corrective maintenance is performed. At this time, update the expected corrective maintenance cost $CMC_s = CMC_s + C_{CM}^{mj}$ and the deterioration amount of the machine $X_{smj} = 0$, and then go to Step 6.2. Otherwise, go to Step 5.

Step 5. Conduct PM and calculate related expected costs. Based on the comparison result of the real-time defective rate $\tilde{p}_{smj}(t)$ of the machine and the quality control threshold QT , decide whether to perform PM. If $\tilde{p}_{smj}(t) \geq QT$, perform maintenance on machine M_{smj} , update the expected PM cost $PMC_s = PMC_s + C_{PM}^{mj}$, the deterioration amount of the machine $X_{smj} = 0$, and the shape parameter of the machine deterioration process $\alpha'_{smj} = \alpha_{smj} \cdot a_{smj}$, and then go to Step 6.1. Otherwise, go to Step 6.2.

Step 6. Conduct OM and calculate related expected costs.

Step 6.1 When a specific machine undergoes PM, determine whether the real-time defective rate of other machine exceeds the OM threshold ω_{smj} . If $QT > \tilde{p}_{smj}(t) \geq \omega_{smj}$, perform OM on the machine, update the expected OM cost $OMC_s = OMC_s + C_{OM}^{mj}$, the deterioration amount of the machine $X_{smj} = 0$, and the shape parameter of the machine deterioration process $\alpha'_{smj} = \alpha_{smj} \cdot a_{smj}$.

Step 6.2 Determine whether the current production order has been completed. If $t \geq I_s$, proceed to Step 7; otherwise, proceed to Step 2.

Step 7. Perform overhaul and calculate related expected costs.

Step 7.1 Determine whether the predicted reliability $R_{smj}^{(s+1)}(I_s | x_{smj})$ of machine M_{smj} is lower than the overhaul threshold ψ_{smj} . If $R_{smj}^{(s+1)}(I_s | x_{smj}) < \psi_{smj}$, perform the overhaul, update the overhaul expected cost $OHC_s = OHC_s + C_{OH}^{mj}$, and set machine degradation $X_{smj} = 0$. Simultaneously, generate a random overhaul time t_{smj}^{ohm} , update the total overhaul duration $T_s^{ohm} = T_s^{ohm} + t_{smj}^{ohm}$, and proceed to step 7.2. Otherwise, proceed to Step 6.2.

Step 7.2 If the machine undergoes an overhaul, update the second and third parts of the expected inventory cost. Second part of the expected inventory cost is $IC_s = IC_s + c_h \cdot \int_0^{t_{IC}^1} (S_{smj} - p_{smj} \cdot t) dt$, where t_{IC}^1 represents the duration of this part. If the overhaul time t_{smj}^{ohm} is less than or equal to the consumption time of the safety stock S_{smj}/p_{smj} , then $t_{IC}^1 = t_{smj}^{ohm}$; If the overhaul time t_{smj}^{ohm} exceeds the consumption time of the safety stock S_{smj}/p_{smj} , then $t_{IC}^1 = S_{smj}/p_{smj}$, leading to a stockout and proceeding to step 7.3. Third part of the expected inventory cost is $IC_s = IC_s + c_h \cdot \int_0^{S_{smj}/p'_{smj}} (p'_{smj} \cdot t) dt$, where p'_{smj} represents the new production speed after the machine resumes production. This speed continues until the inventory level reaches the safety stock.

Step 7.3 Update the expected stockout cost $SC_s = SC_s + c_s \cdot \int_0^{t_{smj}^{ohm} - S_{smj}/p_{smj}} (p_{smj} \cdot t) dt$, and proceed to step 7.4.

Step 7.4 Calculate the total expected cost incurred in the current batch $C_{total} = s \cdot C_{set} + C_{in} + DC_s + SC_s + IC_s + OHC_s + PMC_s + OMC_s + CMC_s$. Update the total time $T_{total} = T_{total} + t$. Update the effective time $ET_s = ET_s + \left\{ \left[I_s \cdot P_s - \int_{t_s}^{t_s+I_s} [\tilde{P}_s(t)] dt \right] \cdot I_s \cdot N_{(m,j)} - \sum_{(m,j) \in G_s^{oh}} \left[\int_{t_s}^{t_s+I_s} (t_{smj}^{ohm}) dt \right] \right\} / N_{(m,j)}$. Check if $s = N_s$; if the condition is met, go to step 8; otherwise, go to step 2.1.

Step 8. Output expected cost rate. $C(W, QT, H, SS) = \frac{C_{total}}{T_{total}}$, proceed to step 9.

Step 9: Output expected effective time utilization rate. $RET(W, QT, H, SS) = \frac{\sum_{s=1}^{N_s} ET_s}{T_{total}}$.

4.2 Model solution by genetic algorithm NSGA-II

As an advanced heuristic search technology, NSGA-II

demonstrates significant advantages in addressing multi-objective optimization problems with extensive and complex solution spaces. Therefore, NSGA-II is introduced in this section to obtain the optimal values of decision variables. Unlike conventional NSGA-II (Su and Liu, 2020), we modified certain aspects to enhance the search efficiency for the optimal values of decision variables concerning quality-based multi-level maintenance. More specifically, a comprehensive optimization framework is constructed to solve complex multi-objective optimization problems efficiently and reliably. The details are as follows.

In the aspect of Encoding, we adopt real number encoding to represent the solutions. Real number encoding facilitates searching in a continuous space, which is particularly effective for problems involving continuous variables.

For generation of the initial population, the initial population consists of $npop$ solutions created within the search space through a random generator. Each solution is represented as a real number vector, with its dimension determined by the complexity of the problem.

For a given solution $X = (x_1, x_2, x_3, x_4)$, its fitness values are calculated as follows: $f_1(X) = C(x_1, x_2, x_3, x_4)$, $f_2(X) = RET(x_1, x_2, x_3, x_4)$, where $C(x_1, x_2, x_3, x_4)$ and $RET(x_1, x_2, x_3, x_4)$ are obtained through the Monte Carlo simulation algorithm described in Section 4.1. During the first calculation, we save the values $f_1(X)$ and $f_2(X)$ corresponding to each solution. If it repeats in subsequent iterations, the corresponding value is returned directly to save computing resources.

In the crossover stage of the genetic algorithm, we employ “Roulette Wheel Selection” or a similar probabilistic selection method to choose parent solutions for crossover. This selection process is based on the fitness value of each solution, giving higher fitness solutions a greater likelihood of being selected. The crossover operation is conducted through arithmetic crossover, a widely used method in real number encoding. Arithmetic crossover generates new offspring solutions by combining two parent solutions. Specifically, given two parent solutions, X_1 and X_2 , and a randomly generated weighting factor λ , the new offspring solutions can be generated in the following way: $X'_1 = \lambda \cdot X_1 + (1 - \lambda) \cdot X_2$; $X'_2 = (1 - \lambda) \cdot X_1 + \lambda \cdot X_2$. In this way, we can explore the new solution space while maintaining the characteristics of the parent solutions.

To explore new search spaces and enhance the diversity of the population, uniform mutation is applied to create new solutions. Uniform mutation is suitable for real-number encoding, where one or more components of a solution are randomly selected for mutation based on a specified mutation probability.

4.3 Bi-objective optimization

As stated in Section 3, there are two objectives: minimizing

the expected cost rate and maximizing the efficiency of effective time utilization. To evaluate the optimal set (W, QT, H, SS) based on the Pareto front solution set obtained from NSGA-II, the CRITIC method is introduced to calculate the objective weights. Subsequently, the TOPSIS method is employed to evaluate and rank each solution in the Pareto front solution set. The specific decision-making process is summarized as follows.

(1) There are g solutions in the Pareto front solution set. Taking the bi-objective function as the evaluation index, the evaluation matrix Z can be obtained as $Z = (z_{uv})_{g \times 2} (u = 1, 2, \dots, g; v = 1, 2)$.

(2) Dimensionless processing of the elements of the index matrix:

The dimensionless operation of the first objective can be expressed as

$$(z'_{u1})_{g \times 1} = \frac{\max(z_{u1}) - z_{u1}}{\max(z_{u1}) - \min(z_{u1})} \quad (u = 1, 2, \dots, g). \quad (21)$$

The second objective is a benefit function, so the dimensionless operation of it can be expressed as

$$(z'_{u2})_{g \times 1} = \frac{z_{u1} - \min(z_{u2})}{\max(z_{u2}) - \min(z_{u2})} \quad (u = 1, 2, \dots, g). \quad (22)$$

The dimensionless index matrix is obtained as $Z' = [(z'_{u1})_{g \times 1}, (z'_{u2})_{g \times 1}]_{g \times 2}$.

(3) Calculate the objective weight ω_v of the index by the CRITIC method:

First, calculate the contrast intensity within each index

as $S_v = \sqrt{\frac{\sum_{u=1}^g (z'_{uv} - \bar{z}'_v)^2}{g - 1}}$ where $\bar{z}'_v = \frac{1}{g} \sum_{u=1}^g z'_{uv}$ is the average within each index.

Secondly, calculate the conflict between the indicators as $R_v = \sum_{i=1}^{n_i} (1 - r_{iv})$, where, r_{iv} represents the correlation coefficient between the indicators i and v , n_i represents the number of indicators. According to the Pearson correlation coefficient calculation formula, it can be expressed

as $r_{iv} = \frac{Cov(z'_{ui}, z'_{iv})}{\sigma_{z'_{ui}} \sigma_{z'_{iv}}}$ where $Cov(z'_{ui}, z'_{iv})$ is the covariance between the indicators. $\sigma_{z'_{ui}}, \sigma_{z'_{iv}}$ are the standard deviations of indicators i and v respectively.

Thirdly, calculate the information amount of each indicator as $C_v = S_v \cdot R_v$.

Finally, the objective weight of each indicator can be obtained as $\omega_v = \frac{C_v}{\sum_{v=1}^2 C_v}$.

(4) Calculate the weighted matrix by multiplying the objective weights ω_v with the indicator matrix to obtain the weighted matrix as $Z'' = (z''_{uv})_{g \times 2} (u = 1, 2, \dots, g; v = 1, 2)$, where, $z''_{uv} = z'_{uv} \cdot \omega_v$.

(5) The maximum element in each column of the weighted matrix is taken as the optimal solution Z_v^+ , and the minimum element in each column is taken as the worst solution Z_v^- : $Z_v^+ = \max(z''_{1v}, z''_{2v}, \dots, z''_{gv})$ and

$$Z_v^- = \min(z''_{1v}, z''_{2v}, \dots, z''_{gv}).$$

(6) Calculate the distances D_u^+ and D_u^- between each element z''_{uv} in the weighted matrix and Z_v^+ , Z_v^- :

$$D_u^+ = \sqrt{\sum_{v=1}^2 (Z_v^+ - z''_{uv})^2} \text{ and } D_u^- = \sqrt{\sum_{v=1}^2 (Z_v^- - z''_{uv})^2}.$$

(7) Calculate the relative closeness S_u of the u -th solution in the Pareto front solution set to the optimal level, and sort them in descending order (the larger the S_u , the closer it is to the optimal level) as $S_u = \frac{D_u^-}{D_u^+ + D_u^-}$.

Finally, select an optimal set of decision variables from the Pareto front solution set as $S_i^* = \max(S_1, S_2, \dots, S_g)$. The comprehensive decision set can be represented as (W_i, QT_i, H_i, SS_i) .

5 Case example

5.1 Case description

This section uses an automotive manufacturing system as an example to illustrate the application of the proposed method. Automotive manufacturers must respond quickly to improve market share. Therefore, MSSB systems are prevalent in this sector. Specifically, the production system examined is commonly employed for manufacturing engine blocks. The precision and smoothness of the inner holes in the engine block directly affect the engine's performance. Therefore, strict quality control is necessary during production. The entire production process is divided into three stages. In the first stage, a lathe (M_{11}) is used for turning, making the excess material of the workpiece to be removed. In the second stage, three horizontal boring machines (M_{21} , M_{22} and M_{23}) are employed to bore the prefabricated holes. In the third stage, two broaching machines (M_{31} , M_{32}) are introduced to complete the through-holes. Elements marked a , b , c in Fig. 4 are represent the three types of machines in each stage respectively.

According to data provided by equipment suppliers and historical operational data, the parameters of each machine are shown in Tables 1 and 2. Additionally, it is assumed that the overhaul time t_{smj}^{ohm} of the machines follows an exponential distribution of which characteristic

parameter is $\theta = 5$.

The considered system can manufacture five types of products ($N_s = 5$). $n(s)$ represents the product type in the production order s . Production parameters for different orders are shown in Table 3–7. Due to that product types, order duration, and processing intensity change dynamically within MSSB, production duration l_s follows a uniform distribution $U[6, 14]$.

To balance solution quality and computation time, the sample size for simulation is set at 30,000 based on the generational distance of objective values and their standard deviations. The solution is represented by real number encoding. The population size is 36, the crossover probability is 0.8, and the mutation probability is 0.05. Figure 5 and Fig. 6 show the results of NSGA-II. Objective values are $C(W^*, QT^*, H^*, SS^*) = 1072$ yuan/day and $RET(W^*, QT^*, H^*, SS^*) = 0.9969$. Optimal values of decision variables are $W^* = 0.04$, $QT^* = 0.015$, $H^* = 0.075$ and $SS^* = 37$ pieces. (The average values are $\bar{C} = 1103.95$ yuan/day, $\overline{RET} = 0.9969$).

5.2 Comparison analysis

To demonstrate the effectiveness of the proposed algorithm, multi-objective particle swarm optimization (MOPSO) is used for comparison analysis. This method is widely applied across various fields. The convergence trends of MOPSO under 40,000 iterations are shown in Figs. 7 and 8. The optimal values can be obtained as $C(W^*, QT^*, H^*, SS^*) = 1001$ yuan/day and $RET(W^*, QT^*, H^*, SS^*) = 0.99212$, with a composite value of 0.98737 after normalizing and multiplying by the weights of the two objectives. Optimal values of decision variables are $W^* = 0.01$, $QT^* = 0.012$, $H^* = 0.0955$ and $SS^* = 105$ pieces.

From Figs. 5 and 6, the average generational distance of the Pareto optimal solution set is approximately 0.42, with a maximum value around 0.8. At the 5000th generation, the standard deviation of the average generational distance decreases to about 0.05, after which it fluctuates around 0.05. In contrast, Figs. 7 and 8 indicate that the average generational distance of the Pareto optimal solution set is around 0.7, with a maximum value of 1. For

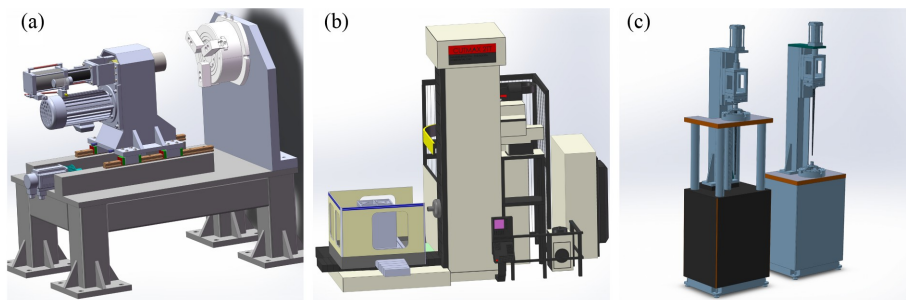


Fig. 4 Machines involved in the case example.

Table 1 Parameter values of the six machines

Parameters	$M_{.11}$	$M_{.21}$	$M_{.22}$	$M_{.23}$	$M_{.31}$	$M_{.32}$	Source
$L_{.mj}$	8.6	7.2	7.6	6.9	10.3	9.8	
α_{smj}	0.38	0.47	0.51	0.42	0.63	0.59	
β_{smj}	0.67	0.73	0.76	0.69	0.87	0.81	
a_{smj}	1.05	1.06	1.08	1.04	1.07	1.06	
$IB_{.mj}$	0.656	0.094	0.094	0.094	0.219	0.219	
\tilde{p}_{0mj}	0.004	0.005	0.005	0.003	0.004	0.004	Cheng and Li (2020)
η_{smj}	0.08	0.07	0.08	0.09	0.06	0.07	
λ_{smj}	0.005	0.004	0.006	0.005	0.004	0.005	
γ_{smj}	1.16	1.2	1.18	1.23	1.27	1.24	
C_{OH}^{mj}	1260	950	860	790	1030	880	
C_{PM}^{mj}, C_{OM}^{mj}	2840	2190	1830	1610	2450	1860	
C_{CM}^{mj}	3280	2570	2190	2080	2910	2220	
b_{1mj}	0.7	0.64	0.51	0.66	0.8	0.84	Assumption
b_{2mj}	0.9	0.73	0.77	1.03	0.93	0.85	

Table 2 Parameters values of the whole system (Source: Cheng and Li, 2020)

Parameters	C_{set}	C_{in}	c_d	c_h	c_s
Values	800	450	65	0.6	80

Table 3 Capacity of six machines in production order s

$n(s)$	p_{s11}	p_{s21}	p_{s22}	p_{s23}	p_{s31}	p_{s32}
1	360	126	90	144	198	162
2	260	87	63	110	140	120
3	240	80	65	95	130	110
4	340	115	85	140	190	150
5	380	135	95	150	210	170

Table 4 Parameter values related to manufacturing process of machines

$n(s)$	d_{s11}	d_{s21}	d_{s22}	d_{s23}	d_{s31}	d_{s32}
1	0.3	0.2	0.5	0.3	0.1	0.3
2	-0.3	-0.6	0.8	0.1	0.7	0.4
3	0.8	0.8	-0.6	-0.5	0.8	-0.4
4	-0.8	-0.8	0.8	0.3	0.3	0.2
5	0.1	-0.3	0.1	0.3	-0.5	0.3

the remaining iterations, the standard deviation fluctuates around 0.04. Although NSGA-II and MOPSO exhibit similar convergence speeds at the 5000th iteration, the stable value of the standard deviation of the average generational distance for MOPSO is lower than that for NSGA-II. Furthermore, the average generational distance of NSGA-II is also lower than that of the MOPSO algorithm. Compared with MOPSO, the average generational distance obtained by NSGA-II is reduced by approximately

Table 5 Parameter values related to processing intensity of machines

$n(s)$	q_{s11}	q_{s21}	q_{s22}	q_{s23}	q_{s31}	q_{s32}
1	0.20	0.07	0.05	0.08	0.11	0.09
2	-0.30	-0.10	-0.07	-0.13	-0.16	-0.14
3	-0.40	-0.13	-0.11	-0.16	-0.22	-0.18
4	0.10	0.03	0.03	0.04	0.06	0.04
5	0.30	0.11	0.08	0.12	0.17	0.13

Table 6 Parameter values related to processing intensity of machines after overhaul

$n(s)$	q'_{s11}	q'_{s21}	q'_{s22}	q'_{s23}	q'_{s31}	q'_{s32}
1	0.40	0.14	0.10	0.16	0.22	0.18
2	-0.60	-0.20	-0.15	-0.25	-0.32	-0.28
3	-0.80	-0.27	-0.22	-0.32	-0.43	-0.37
4	0.20	0.07	0.05	0.08	0.11	0.09
5	0.60	0.21	0.15	0.24	0.33	0.27

Table 7 Capacity of six machines in production order s after overhaul

$n(s)$	p'_{s11}	p'_{s21}	p'_{s22}	p'_{s23}	p'_{s31}	p'_{s32}
1	720	252	180	288	396	324
2	520	174	126	220	280	240
3	480	160	130	190	260	220
4	680	230	170	280	380	300
5	760	270	190	300	420	340

$(0.7 - 0.42)/0.7 = 40\%$. This means that result of NSGA-II algorithm is closer to the global optimal value.

Figures 9 and 10 are the final Pareto front lines obtained by NSGA-II and MOPSO, respectively.

Moreover, several typical maintenance policies are

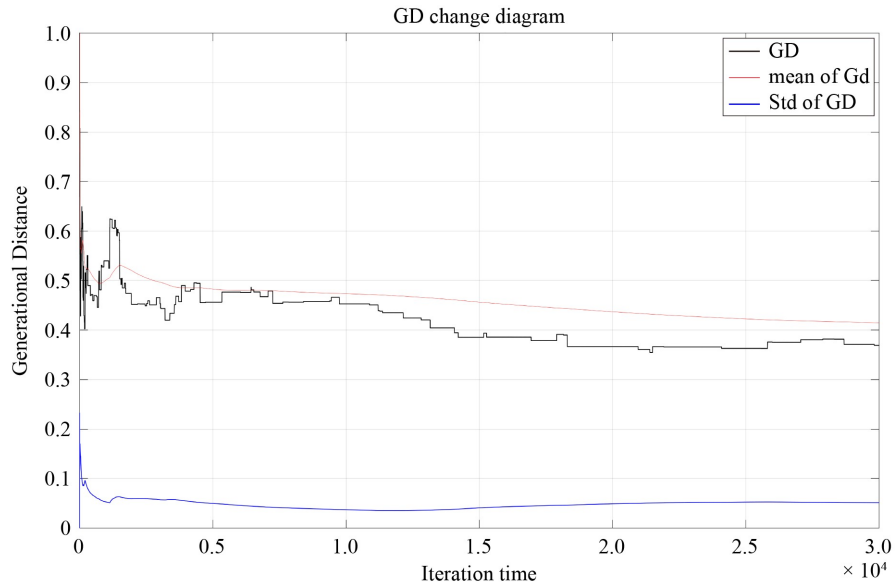


Fig. 5 Generational distance obtained by NSGA-II at each generation.

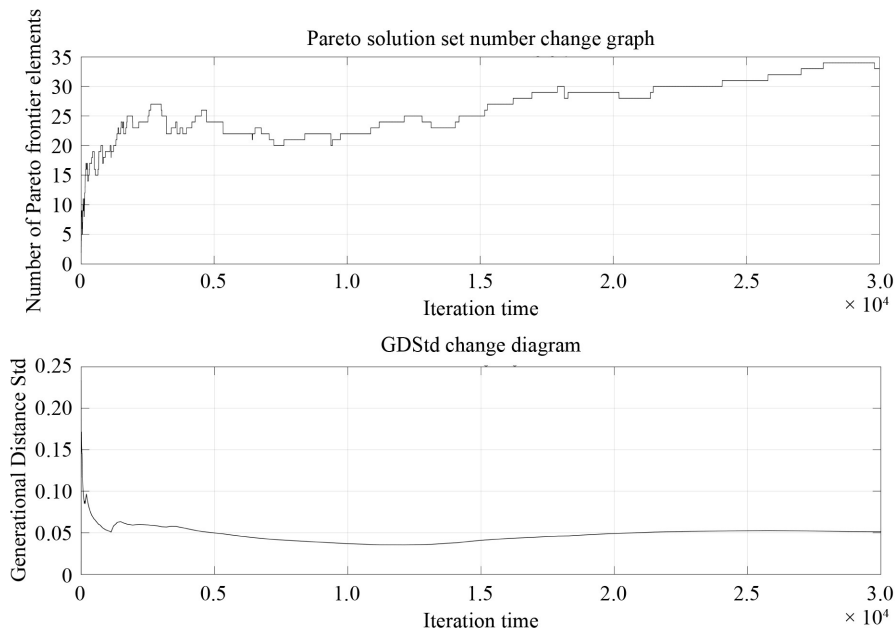


Fig. 6 Convergence of NSGA-II.

introduced for comparison with the proposed maintenance approach. Comparison I refers to a widely applied maintenance policy that disregards buffer stock. This policy has been utilized in numerous existing studies (Zhang et al., 2024; Liu et al., 2018). Comparison II represents the policy that simultaneously optimizes condition-based maintenance and buffer stock, where maintenance thresholds are based on cumulative degradation levels rather than product quality (Dehghan Shoorkand et al., 2024; Zheng et al., 2021). As indicated in Table 8, the proposed policy achieves superior levels in both expected cost rate and effective time rate. First, when buffer stock is

neglected, productivity decreases due to wasted machine repair time. The expected shortage cost increases, which may reduce the effective time rate and elevate the expected cost rate. Furthermore, when maintenance actions are condition-based, inspection downtime for machines increases, adversely affecting both expected cost rate and effective time rate.

5.3 Sensitivity analysis

In this section, sensitivity analysis of decision variables includes W^* , QT^* , H^* and SS^* are conducted to specify

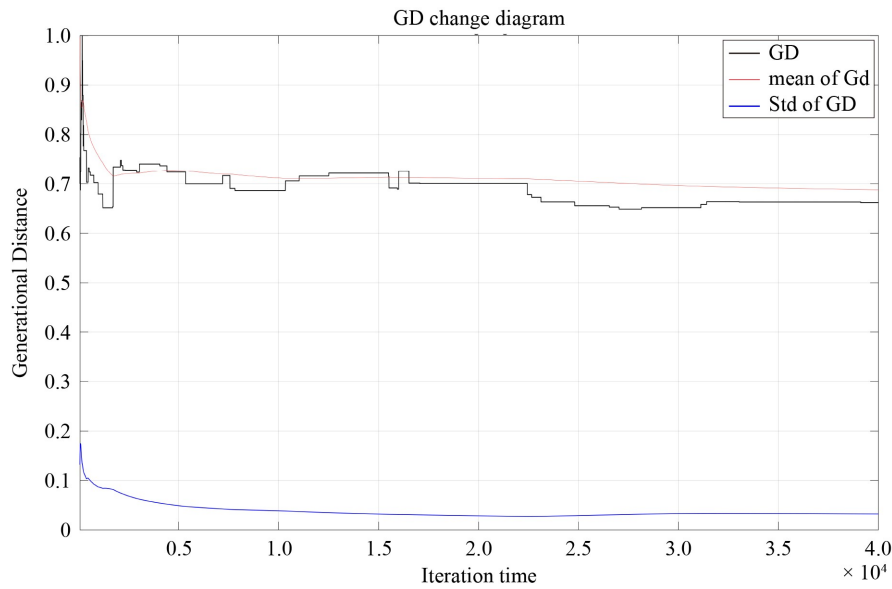


Fig. 7 Generational distance obtained by MOPSO at each generation.

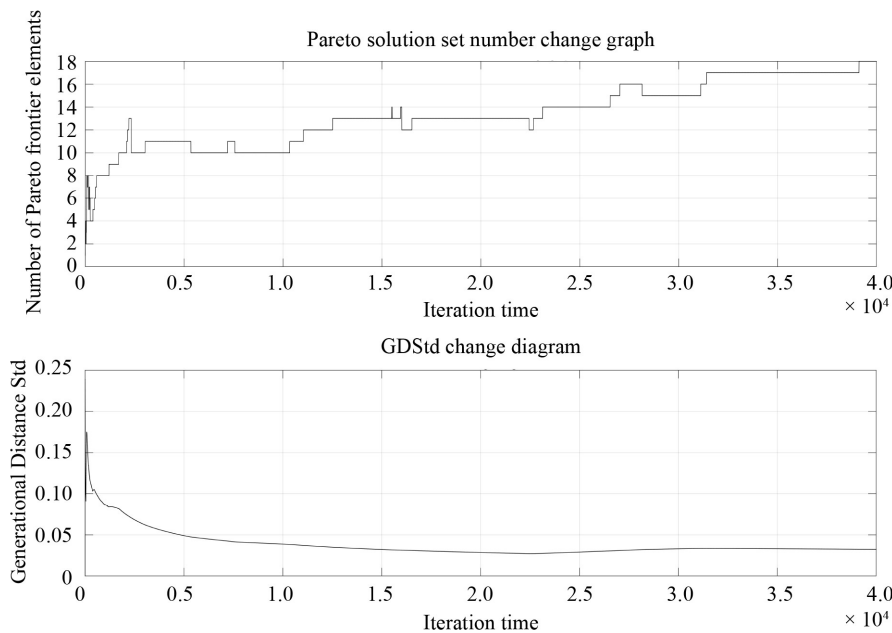


Fig. 8 Convergence of MOPSP.

Table 8 Results obtained under different strategies

policy	$C(W^*, QT^*, H^*, SS^*)$	$RET(W^*, QT^*, H^*, SS^*)$
comparison I	1179	0.9584
comparison II	1225	0.9670
proposed method	1072	0.9969

Table 9 result of sensitivity ranking of decision variables

decision variable	sensitivity ranking
W^*	$C_{set} < C_{PM} < c_h < C_{CM} < C_{in} < c_s < C_{OH} < \theta < c_d$
QT^*	$c_h < \theta < C_{OH} < C_{in} < c_d < C_{set} < C_{CM} < C_{PM} < c_s$
H^*	$c_s < C_{set} < c_h < C_{in} < C_{PM} < c_d < C_{CM} < \theta < C_{OH}$
SS^*	$\theta < c_d < C_{set} < C_{CM} < C_{PM} < c_s < C_{in} < c_h < C_{OH}$

how the optimal values of them are influenced by others considered parameters. These parameters are assumed to vary from -50% to + 50%. Results are illustrated in Figs. 11–14. Based on the variance of the collected decision variable values, the sensitivity ranking of each variable is

summarized in Table 9. The expected cost of overhaul may significantly influence the buffer stock level and OM threshold. The expected costs resulting from buffer shortages and defective products have the most adverse effects

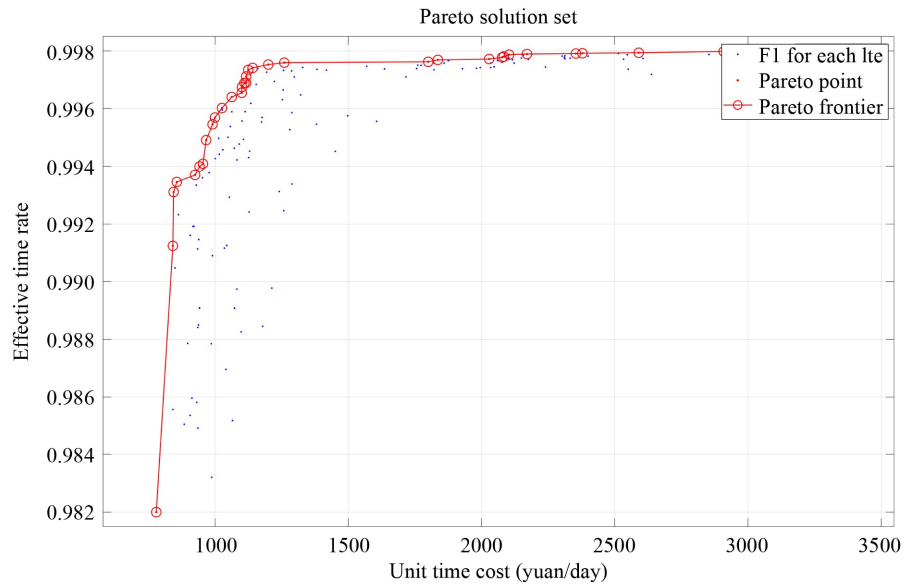


Fig. 9 Pareto frontier obtained by NSGA-II.

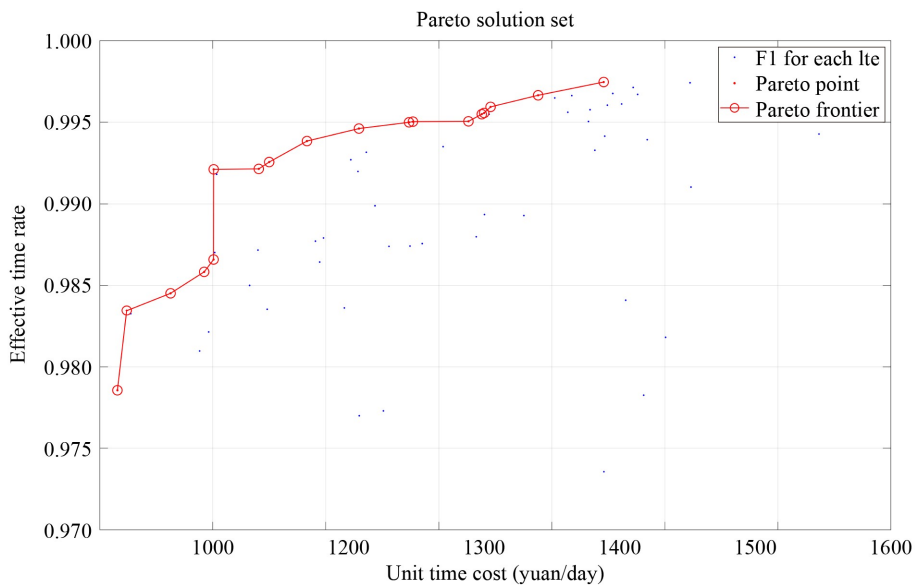


Fig. 10 Pareto frontier obtained by MOPSOC.

on the threshold levels of quality and overhaul.

6 Conclusions

6.1 Theoretical contribution

This paper focuses on multi-stage production systems for the MSSB mode. The influences of different product types on the involved machines are specified. The relationship between machine performance efficiency and the quality level of output products is explored. A buffer

stock policy is proposed to meet customer demand during interruptions caused by maintenance actions. The main contributions of this paper include: 1) A multi-stage Gamma process is developed to analyze the effects of different product specifications on machine reliability within MSSB. 2) A quality-based maintenance policy is designed, allowing for the scheduling of maintenance actions without stopping machines to inspect their performance levels. 3) Buffer stock is optimized alongside product quality and machine maintenance, which can enhance operational management efficiency. 4) A simulation-based NSGA-II is proposed to search for the optimal

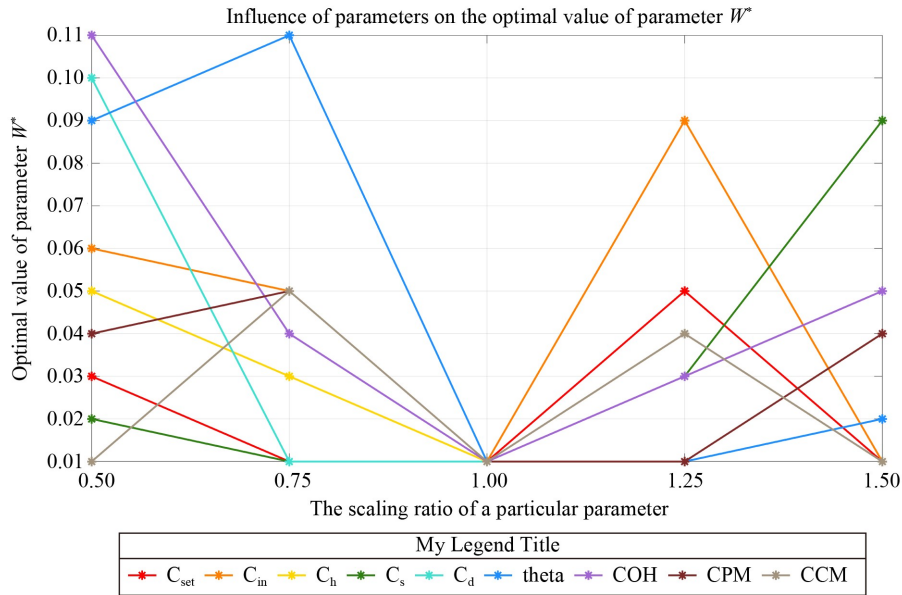


Fig. 11 Influence of parameters on the optimal value of W^* .

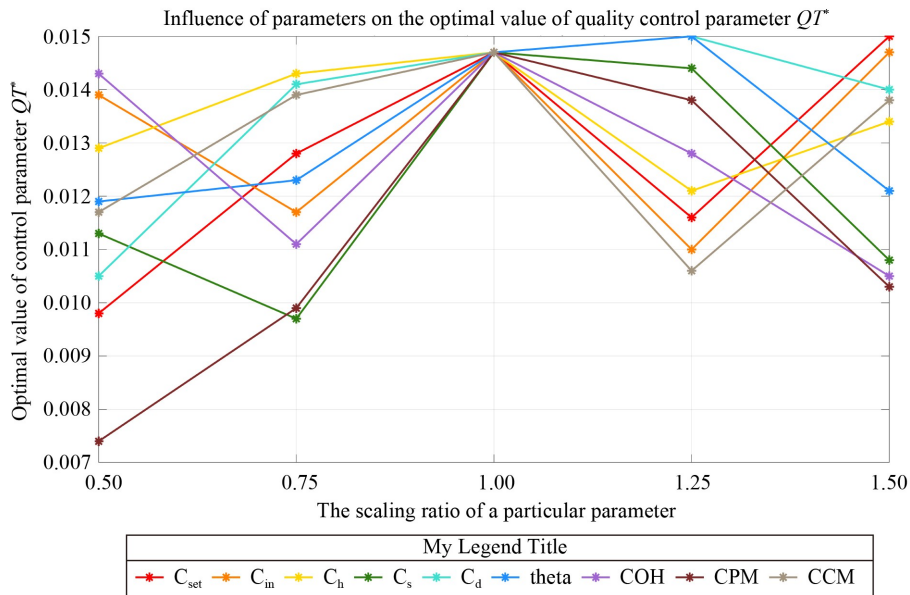


Fig. 12 Influence of parameters on the optimal value of QT^* .

values of decision variables under multiple conflicting objectives.

6.2 Managerial implications

When applying the proposed method in practice, positive effects can be observed in the following areas. (1) The accuracy of machine reliability evaluation can be improved since the effects of different product specifications on machine performance are specified. Operations managers can gain deeper insights into machine perfor-

mance variation, particularly within MSSB, where the types of products processed are multiple and varied. (2) Maintenance actions for production systems can be scheduled at a lower cost. The cost reduction arises from minimizing related activities arranged for performance inspection. By analyzing the relationship between machine degradation levels and product defect rates, maintenance actions can be scheduled in real time, enhancing the agility of operational management. (3) By optimizing the safety stock level alongside maintenance and quality control characteristics, the advantages of the

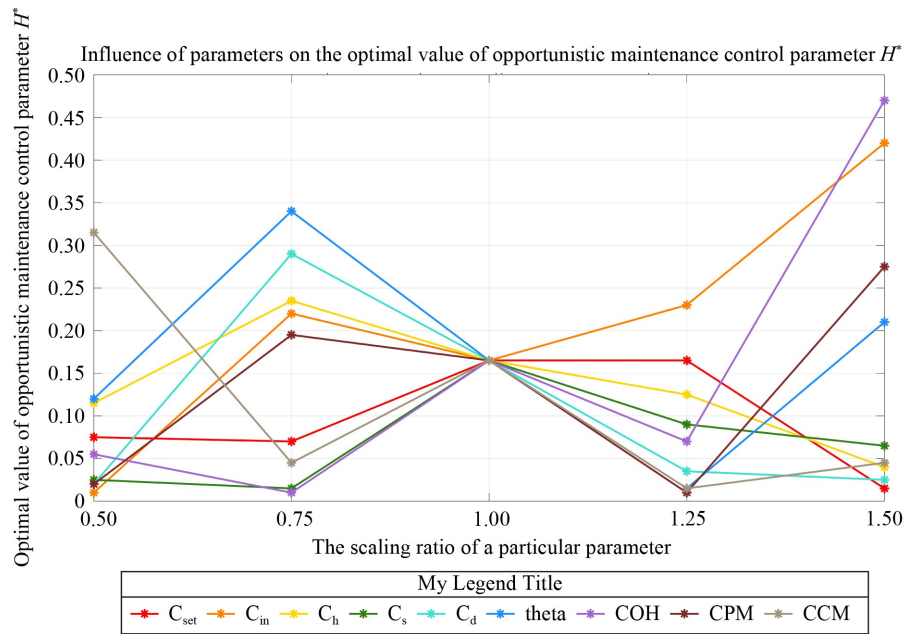


Fig. 13 Influence of parameters on the optimal value of H^* .

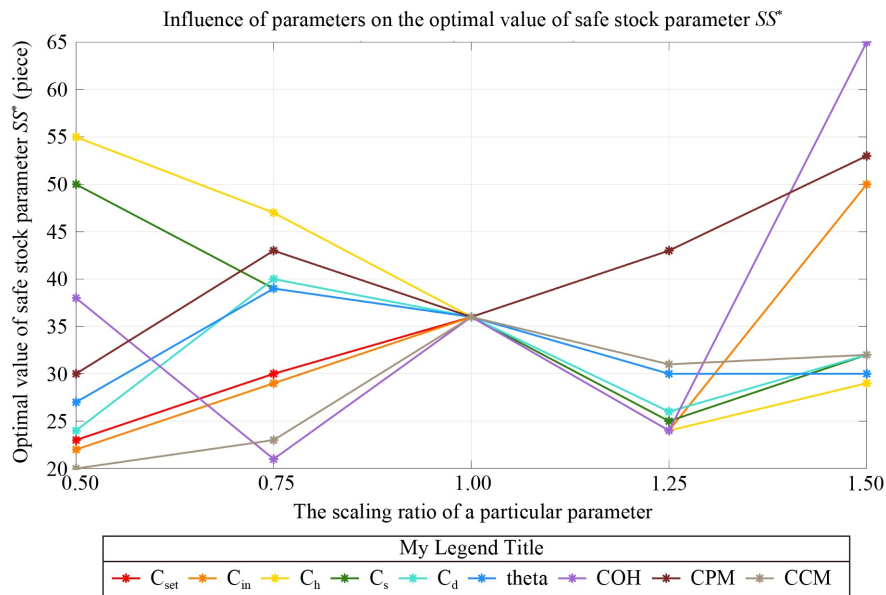


Fig. 14 Influence of parameters on the optimal value of SS^* .

MSSB production model can be maintained.

6.3 Future focus

Although the effects of different product specifications on machine degradation are specified, the failure mechanisms of all machines are assumed to be consistent. This assumption may overlook the differences among machines. For example, some machines degrade continuously, while others may degrade discretely, meaning their

degradation states are limited to a finite set of conditions. Given that the considered production system consists of multiple different machines, relationships among machine performance states, product defect rates, maintenance policies, and buffer stock policies must be modified. These aspects will be the focus of future research following this paper.

Competing Interests The authors declare that they have no competing interests.

Notations

N_s	Total quantity of production orders
l_s	Length of the production cycle of the s th order
q_{smj}	Processing intensity of machine M_{smj} in the s th order
q'_{smj}	Processing intensity of machine M_{smj} after overhaul
b_{1mj}	Coefficient of influence of processing technology on machine M_{smj}
b_{2mj}	Coefficient of influence of processing strength on machine M_{smj}
P_s	System productivity in the s th order
D_s	Market demand rate of the s th order
p_{smj}	Production rate of machine M_{smj} in the s th order
p'_{smj}	Production rate of machine M_{smj} in the s th order after overhaul
IB_{mj}	Structure importance measure of machine M_{smj}
$\bar{p}_{smj}(t)$	Defective rate of machine M_{smj} at time t in the s th order
\bar{p}_{0mj}	Initial defective rate of machine M_{smj} in a brand new state
$\eta_{smj}, \lambda_{smj}, \gamma_{smj}$	Parameters of quality deterioration in the s th order
$\alpha_{smj}, \beta_{smj}$	Shape and scale parameters of gamma process in the s th order
$X_{smj}(t)$	Cumulative gradation level of machine M_{smj} at time t in the s th order
L_{smj}	Failure threshold of machine M_{smj} in the s th order
CR_{smj}	Capacity ratio of equipment M_{smj} in the s th order in the m th stage
t_{smj}^{ohm}	Duration of overhaul performed on machine M_{smj} in the s th order
a_{smj}	ADGP parameter of machine M_{smj} in the s th order
C_{set}	Production line replacement cost (preparation cost of production order)
C_{in}	Cost of inspection for the system
c_d	Cost of loss of defective products per unit item
c_h	Inventory cost per unit item per unit time
c_s	Shortage cost of an item per unit time
C_{PM}^{mj}	Preventive maintenance cost for machine M_{smj}
C_{OM}^{mj}	Opportunistic maintenance cost for machine M_{smj}
C_{CM}^{mj}	Corrective maintenance cost for machine M_{smj}
C_{OH}^{mj}	Overhaul cost for machine M_{smj}
ET_s	Effective time rate of the s th order
g, u, v	Parameters of evaluation matrix Z
S_u	The relative proximity of the solutions in the Pareto frontier solution set

References

Azimpoor S, Taghipour S (2021). Joint inspection and product quality optimization for a system with delayed failure. *Reliability Engineering & System Safety*, 215: 107793

Bevilacqua M, Braglia M (2000). The analytic hierarchy process applied to maintenance strategy selection. *Reliability Engineering*

& System Safety, 70(1): 71–83

Boumallessa Z, Chouikhi H, Elleuch M, Bentaher H (2023). Modeling and optimizing the maintenance schedule using dynamic quality and machine condition monitors in an unreliable single production system. *Reliability Engineering & System Safety*, 235: 109216

Bouslah B, Gharbi A, Pellerin R (2016). Joint economic design of production, continuous sampling inspection and preventive maintenance of a deteriorating production system. *International Journal of Production Economics*, 173: 184–198

Cao Y, Wang P, Xv W, Dong W (2024). Criticality analysis for continuous degrading systems subject to multi-level failure dependences. *Computers & Industrial Engineering*, 194: 110395

Chen Z, Chen Z, Zhou D, Xia T, Pan E (2023). Opportunistic maintenance optimization of continuous process manufacturing systems considering imperfect maintenance with epistemic uncertainty. *Journal of Manufacturing Systems*, 71: 406–420

Cheng G, Li L (2020). Joint optimization of production, quality control and maintenance for serial-parallel multistage production systems. *Reliability Engineering & System Safety*, 204: 107146

Cheng G, Zhou B, Li L (2018). Integrated production, quality control and condition-based maintenance for imperfect production systems. *Reliability Engineering & System Safety*, 175: 251–264

Chiu Y, Chen K, Ting C (2012). Replenishment run time problem with machine breakdown and failure in rework. *Expert Systems with Applications*, 39(1): 1291–1297

de Jonge B, Scarf P A (2020). A review on maintenance optimization. *European Journal of Operational Research*, 285(3): 805–824

Dehghan Shoorkand H, Nourelfath M, Hajji A (2024). A hybrid deep learning approach to integrate predictive maintenance and production planning for multi-state systems. *Journal of Manufacturing Systems*, 74: 397–410

Gan J, Zhang W, Wang S, Zhang X (2022). Joint decision of condition-based opportunistic maintenance and scheduling for multi-component production systems. *International Journal of Production Research*, 60(17): 5155–5175

Hadian S, Farughi H, Rasay H (2021). Joint planning of maintenance, buffer stock and quality control for unreliable, imperfect manufacturing systems. *Computers & Industrial Engineering*, 157: 107304

Hu J, Shen J, Shen L (2020). Periodic preventive maintenance planning for systems working under a Markovian operating condition. *Computers & Industrial Engineering*, 142: 106291

Hu J, Sun Q, Ye Z S (2022). Replacement and repair optimization for production systems under random production waits. *IEEE Transactions on Reliability*, 71(4): 1488–1500

Jain A K, Lad B K (2017). Dynamic optimization of process quality control and maintenance planning. *IEEE Transactions on Reliability*, 66(2): 502–517

Kuo W, Zhu X (2012). Some recent advances on importance measures in reliability. *IEEE Transactions on Reliability*, 61(2): 344–360

Liu Q, Dong M, Chen F (2018). Single-machine-based joint optimization of predictive maintenance planning and production scheduling. *Robotics and Computer-integrated Manufacturing*, 51: 238–247

Liu Q, Dong M, Frank Chen F, Liu W, Ye C (2020). Multi-objective imperfect maintenance optimization for production system with an intermediate buffer. *Journal of Manufacturing Systems*, 56: 452–462

- Lopes R (2018). Integrated model of quality inspection, preventive maintenance and buffer stock in an imperfect production system. *Computers & Industrial Engineering*, 126: 650–656
- Mohammad Hadian S, Farughi H, Rasay H (2023). Development of a simulation-based optimization approach to integrate the decisions of maintenance planning and safety stock determination in deteriorating manufacturing systems. *Computers & Industrial Engineering*, 178: 109132
- Polotski V, Gharbi A, Kenne J P (2022). Production control in manufacturing systems with perishable products under periodic demand. *Journal of Manufacturing Systems*, 63: 288–303
- Shi H, Zhang J, Zio E, Zhao X (2023). Opportunistic maintenance policies for multi-machine production systems with quality and availability improvement. *Reliability Engineering & System Safety*, 234: 109183
- Su C, Liu Y (2020). Multi-objective imperfect preventive maintenance optimisation with NSGA-II. *International Journal of Production Research*, 58(13): 4033–4049
- Sun Q, Chen P, Wang X, Ye Z S (2023). Robust condition-based production and maintenance planning for degradation management. *Production and Operations Management*, 32(12): 3951–3967
- Tasias K (2022). Integrated quality, maintenance and production model for multivariate processes: A Bayesian approach. *Journal of Manufacturing Systems*, 63: 35–51
- Wan Q, Chen L, Zhu M (2023). A reliability-oriented integration model of production control, adaptive quality control policy and maintenance planning for continuous flow processes. *Computers & Industrial Engineering*, 176: 108985
- Wang K, Jing H, Wang D, Jiang F (2024). Joint quality and maintenance decisions under servitization business model. *International Journal of Production Research*, 62(5): 1567–1585
- Zhang N, Cai K, Deng Y, Zhang J (2024). Joint optimization of condition-based maintenance and condition-based production of a single equipment considering random yield and maintenance delay. *Reliability Engineering & System Safety*, 241: 109694
- Zhang Z, Yang L (2021). State-based opportunistic maintenance with multifunctional maintenance windows. *IEEE Transactions on Reliability*, 70(4): 1481–1494
- Zheng R, Zhou Y, Gu L, Zhang Z (2021). Joint optimization of lot sizing and condition-based maintenance for a production system using the proportional hazards model. *Computers & Industrial Engineering*, 154: 107157
- Zhou X, Ning X (2021). Maintenance gravity window based opportunistic maintenance scheduling for multi-unit serial systems with stochastic production waits. *Reliability Engineering & System Safety*, 215: 107828
- Zhou X, Yu M (2020). Semi-dynamic maintenance scheduling for multi-station series systems in multi-specification and small-batch production. *Reliability Engineering & System Safety*, 195: 106753
- Zhou X, Zhu M, Yu M (2021). Maintenance scheduling for flexible multistage manufacturing systems with uncertain demands. *International Journal of Production Research*, 59(19): 5831–5843
- Zhu M, Zhou X (2022). Lifecycle maintenance scheduling of station with multiple categories of information. *Computers & Industrial Engineering*, 172: 108593
- Zhu M, Zhou X (2023). Hierarchical-clustering-based joint optimization of spare part provision and maintenance scheduling for serial-parallel multi-station manufacturing systems. *International Journal of Production Economics*, 264: 108971
- Zhu X, Wang J, Coit D W (2024). Joint optimization of spare part supply and opportunistic condition-based maintenance for onshore wind farms considering maintenance route. *IEEE Transactions on Engineering Management*, 71: 1086–1102