RESEARCH ARTICLE

Optimization and Control of Discrete-Time Production-Inventory Systems Using Reinforcement Learning

Renfang Wang[1], **Yufei Gong**[1], **Peng Su**[2], **Linmin Hu**[1,*] **and Xin Jiang**[1]

[1] School of Science, Yanshan University, Qinhuangdao 066004, China
[2] School of Economics and Management, North University of China, Taiyuan 030051, China

Abstract

This study introduces a novel approach for enhancing production decision-making by applying Reinforcement Learning to optimize the Economic Manufacturing Quantity (EMQ) model within discrete-time production-inventory systems. By incorporating machine status, inventory levels, and production choices, a Markov Decision Process (MDP) is constructed and combined with the Q-learning algorithm to derive an adaptive control method. This method enables the dynamic adaptation of production decisions, by effectively balancing the normal operation and shutdown for rest states. Numerical simulations show that the suggested Reinforcement Learning model surpasses conventional EMQ models and steady-state probability models in both convergence speed and cost-effectiveness. This study offers a data-driven approach for optimizing production processes in smart manufacturing settings. It also supports the evolution of production-inventory systems from static planning to dynamic intelligent decision-making.

Keywords: reinforcement learning, economic manufacturing quantity, production inventory optimization, Q-Learning, dynamic decision-making.

1 Introduction

The advancement of global manufacturing towards intelligence and digitalization presents unprecedented challenges in complexity and uncertainty for the operation management of production systems. Traditional production inventory management theories rely heavily on static Economic Manufacturing Quantity (EMQ) models. While effective in deterministic environments, these models struggle to address dynamic uncertainties like random machine failures, market demand fluctuations, and supply chain disruptions in real-world production settings. With the rise of personalized customization and flexible manufacturing, enterprises now require production systems with enhanced real-time responsiveness and adaptive optimization capabilities. Therefore, the establishment of intelligent and flexible production inventory control systems has become a critical concern for modern manufacturing enterprises aiming to boost operational efficiency and cut production costs.

Submitted: 22 August 2025
Accepted: 22 September 2025
Published: 11 November 2025

Vol. 1, **No.** 2, 2025.
10.62762/TSSR.2025.621059

***Corresponding author:**
✉ Linmin Hu
linminhu@ysu.edu.cn

Citation
Wang, R., Gong, Y., Su, P., Hu, L., & Jiang, X. (2025). Optimization and Control of Discrete-Time Production-Inventory Systems Using Reinforcement Learning. *ICCK Transactions on Systems Safety and Reliability*, 1(2), 98–113.

© 2025 ICCK (Institute of Central Computation and Knowledge)

A typical industrial application of the classical EMQ model can be illustrated with the production of a specific type of small copper connector. The monthly demand for this product is 1,230 units, and the unit production cost is $0.0135. However, the setup cost per production run is relatively high, at $2.15. Based on the EMQ formula, the optimal production batch size is approximately 6,850 units—enough to cover about six months of demand [1]. This example demonstrates the ability of traditional EMQ models to effectively balance setup and inventory holding costs. At the same time, it reveals inherent limitations: excessively small batches increase setup frequency, reducing equipment utilization and production efficiency; conversely, overly large batches result in inventory accumulation and higher risks of material deterioration. Moreover, when faced with uncertainties such as stochastic machine failures or demand variability, traditional static models like the EMQ lack the flexibility to support dynamic optimization.

The swift advancement of artificial intelligence technology has propelled Reinforcement Learning (RL) to a leading optimization algorithm. This algorithm, as a machine learning approach, acquires optimal decision-making skills by engaging with the environment. This methodology offers innovative solutions for managing dynamic decision-making challenges within production inventory systems. In this domain, the value function-centered Q-learning algorithm has surfaced as a potent technique for resolving discrete state-space decision predicaments, owing to its straightforward implementation and advantageous convergence characteristics.

This study is grounded in discrete-time production systems, integrating Reinforcement Learning theory into traditional production and inventory management. The aim is to create an intelligent control framework that integrates machine state, inventory levels, and production decisions. To address the shortcomings of conventional EMQ models in handling stochastic machine failures, a MDP model is developed, taking into account machine failures and repair times. An adaptive decision-making strategy is proposed, employing the Q-learning algorithm to continuously optimize production strategies and monitor operational costs.

2 Literature Review

The escalating intricacy of global supply chains and heightened market competitiveness have advanced discrete-time inventory systems research to the forefront of operations management. Advancements in modeling and optimizing of these systems primarily concentrate on fundamental aspects, including demand patterns modeling, order policies optimization and costs controlling.

2.1 Traditional EMQ Model

The EMQ model has undergone continuous refinement and development within a discrete-time framework, with researchers expanding its applicability by introducing new constraints and optimization objectives. Jaber et al. [2] systematically analyzed the influence of learning curves on the Economic Production/Order Quantity model, investigating the integration of just-in-time production principles. Giri et al. [3] formulated a discrete-time EMQ model incorporating geometric discrete failure and repair time distributions, determined the optimal production time criterion under cost minimization. Giri et al. [4] introduced the production rate as a decision variable in the EMQ model, established a dynamic optimization framework where production rate is contingent on failure rates and costs, and thereby achieves a balance between service levels and costs through safety stock strategies. Chiu et al. [5] optimized an EMQ model by considering defective item rework and multi-batch shipments through mathematical modeling, and derived closed-form solutions for optimal production batch sizes and costs. Chiu et al. [6] proposed an EMQ model on the basis of the Bass diffusion model for addressing the time-varying dynamical demands throughout a product's lifecycle. Sarkar et al. [7] developed an EMQ model that integrates price- and time-dependent demand, product reliability, and inflation, effectively integrates dynamic cost considerations and defective item rework.

2.2 Joint Optimization of Maintenance Strategy and Inventory Management

The integration of predictive maintenance technology has sparked significant interest in the collaborative optimization of production planning and equipment maintenance. Borrero et al. [8] utilized MDP to develop a dynamic maintenance strategy that considers machine age and inventory level. This strategy, tailored for single-machine, single-product scenarios, yielded substantial cost reductions comparing to conventional static approaches. Zhang et al. [9] introduced an integrated model that combines EMQ with condition-based maintenance, addressed joint optimization issues of production

and maintenance in the presence of imperfect manufacturing processes and inspection errors. Building on this work, Zhang et al. [10] formulated an MDP model for the joint optimization of production and maintenance by modeling equipment degradation as a discrete-time Markov chain within the framework of condition-based maintenance and production. Han et al. [11] merged condition monitoring with the Wiener process to devise an optimization model for continuous production systems. This model incorporates EMQ and hybrid maintenance strategies to determine the expected cost rate and optimize the production batch sizes. Tan et al. [12] assumed exponential distributions for demand arriving time, production time, and warm-up time, employs MDP and linear programming to derive an optimal control strategy. Pazouki et al. [13] proposed a dynamic manufacturing inventory system model that accounts for varying demand and return rates to maximize profits and, showcased the economic and environmental benefits of high-quality production. Li et al. [14] developed a production-inventory system that incorporates the (s, S) replenishment policy based on a Markov chain framework. They identified the optimal replenishment strategy by evaluating system metrics and utilizing the Non-dominated Sorting Genetic Algorithm.

2.3 Data-Driven and Reinforcement Learning

In recent years, RL methods have offered innovative solutions for complex inventory decision-making problems. Wu et al. [15] introduced a new deep Reinforcement Learning method. This method uses derivative-free optimization to solve inventory management problems in multi-echelon supply chains. Hubert et al. [16] successfully optimized inventory production scheduling through the integration of Deep Reinforcement Learning with discrete event simulation. Zhou et al. [17] developed a multi-echelon spare parts inventory optimization model based on Multi-Agent Deep Reinforcement Learning. This model combines value decomposition with the twin delayed deep deterministic policy gradient algorithm to dynamically adjust inventory policies and address coordination challenges among multiple warehouses. Tian et al. [18] introduced a Deep Reinforcement Learning framework that integrates the advantage Actor-Critic algorithm with Proximal Policy Optimization, demonstrating significant reductions in inventory costs compared to single-algorithm approaches.

Although the existing literature has conducted in-depth research on traditional EMQ models, inventory management, and production strategies, most studies do not fully account for the interaction mechanisms between dynamic changes in machine states and production decisions. In complex environments characterized by random faults and multiple states, traditional methods often rely on strong assumptions or static strategies, which complicates the achievement of dynamic optimization. The absence of an intelligent decision-making framework capable of autonomously learning and adjusting production strategies during operations results in a significant gap between theoretical models and actual production environments. This paper innovatively integrates reinforcement learning theory into the EMQ model to construct a joint optimization framework that encompasses machine state, inventory level, and production actions. By employing a Q-learning algorithm, it facilitates state-dependent adaptive decision-making and provides a more accurate representation of how random machine failures, repairs, and dynamic inventory changes impact production strategy. Consequently, this approach enables optimal production cost reduction through flexible adjustments to production strategies within highly uncertain environments.

The remainder of this paper is structured as follows. Section 3 introduces the fundamental notations of the model and outlines the essential assumptions. Section 4 develops three distinct types of production-inventory models and derives three critical metrics: the average total cost per time unit, the average production-inventory cycle, and the average total cost per cycle. Section 5 presents numerical experiments that compare scenarios with and without maintenance cost during machine shutdown for rest, thereby validates the model's effectiveness through key performance indicators. Section 6 concludes the paper.

3 Symbol Explanation and Model Assumptions

3.1 Symbol Explanation

- t: Discrete times, $t = 0, 1, 2, \ldots$
- M: Machine status, $M = 0$ (complete failure), $M = 1$ (normal operation), $M = 2$ (shutdown for rest)
- I: Inventory level, $I = 0, 1, 2, \ldots, I_{\max}$

- S: System status, $S = (M, I)$
- A: Action space, $A = \{0, 1\}$, where 0 represents no production and 1 represents production
- a: The action selected by machine during the operation, $a \in A$
- $\boldsymbol{\pi}$: Steady-state probability distribution vector under long-term system operation
- p: Failure rate of machine per time unit
- q: Repair rate of machine per time unit
- u: Production rate of machine per time unit
- d: Demand rate of machine per time unit
- C_p: Single production cost
- C_h: Inventory holding cost per unit of product per unit of time
- C_b: Single stockout penalty cost
- C_f: Maintenance cost of machine per time unit
- C_r: Maintenance cost during machine shutdown for rest
- $C(S_t, a_t)$: The immediate cost at time t
- T: The average production-inventory cycle
- U: The average total cost per cycle
- ETU: The average total cost per unit of time
- T_I, T_{II}, T_{III}: The average production-inventory cycles corresponding to Model I, II, and III, respectively
- U_I, U_{II}, U_{III}: The average total cost per cycle corresponding to Model I, II, and III, respectively
- T_k: The time interval of the k-th cycle in Model III
- U_k: The total cost of the k-th cycle in Model III
- $ETU_I, ETU_{II}, ETU_{III}$: The average total cost per unit of time corresponding to Model I, II, and III, respectively

3.2 Model Assumptions

This paper analyzes the single-machine production-inventory system characterized by three distinct states: complete failure, normal operation, and shutdown for rest. The durations of failures and repairs are modeled using geometric distributions. By integrating the machine's state with inventory levels, we establish MDP. A Reinforcement Learning model is employed to optimize the production strategy, with the objective of minimizing long-term operational costs. To enable a more detailed examination of the system, we make the following assumptions:

Assumption 1: u/d is an integer greater than 1, i.e., $u = (t + 1)d$, where $t = 1, 2, 3....$

Assumption 2: Both machine failures and repair times are modeled using geometric distributions. Upon the occurrence of a failure, repair work commences immediately. Once the repair is completed, the machine transitions to the load operation state only when $I = 0$; otherwise, it enters a shutdown for rest state.

Assumption 3: The state of machine operation and the changes of inventory level occur at the end of discrete time.

Assumption 4: When the inventory reaches the maximum level I_{max}, the machine is forced to enter a shutdown for rest state.

Assumption 5: The single production cost only takes effect when $a = 1$.

Assumption 6: During instances of complete machine failure and shutdown for rest periods, the demand is satisfied by the accumulated inventories. Any unmet demand resulting from extended maintenance durations is regarded as entirely lost (see Figure 1).

Assumption 7: When the inventory level $I = 0$, machine state $M = 1$ and action state $a = 1$, a new production cycle is started (see Figure 1).

4 Model Design and Construction

To systematically analyze the cost optimization problem in production-inventory systems, this section establishes two benchmark models based on conventional methodologies and employs analytical approaches to determine the long-term operating costs and cycles of the system. Subsequently, a dynamic decision-making model grounded in Reinforcement Learning is introduced, which optimizes dynamic production strategies through interactions between agents and their environment.

4.1 EMQ Model in Discrete Time (Model I)

Giri et al. [3] conducted an investigation into the EMQ model under discrete-time conditions, taking into account stochastic machine failures and repairs. According to Figure 1, the cycle period is defined as the time interval between two consecutive production

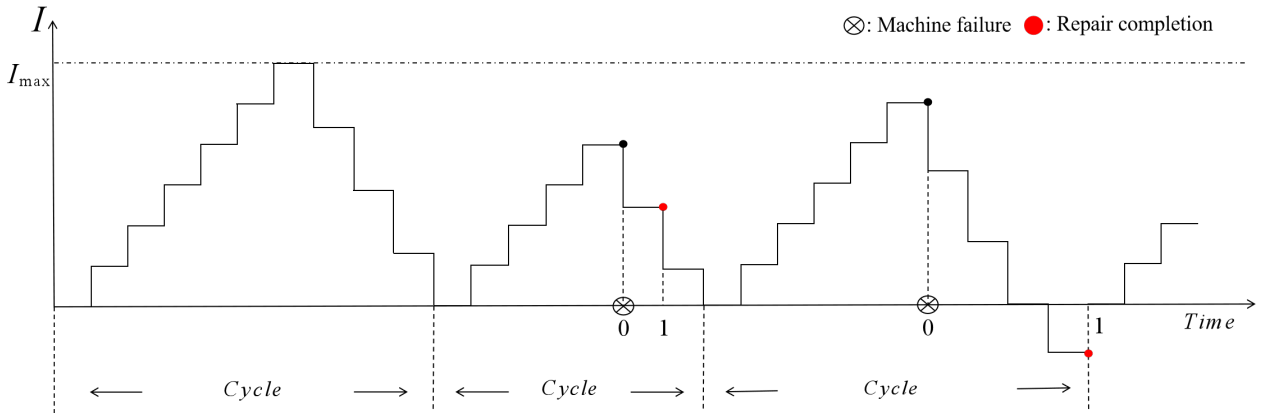

Figure 1. Changes in inventory levels.

points. Subsequently, the average total cost per cycle and the average inventory holding period during machine failures and repair times—when these durations follow a general distribution—are derived using discrete probability theory:

$$\begin{aligned} U(t_0) = {} & C_p + C_f \sum_{t=0}^{t_0-1} \sum_{\tau=0}^{\infty} \tau g(\tau) p(t) + C_h \sum_{t=0}^{t_0-1} \frac{u(u-d)}{2d} t^2 p(t) \\ & + C_h \sum_{t=t_0}^{\infty} \frac{u(u-d)}{2d} t_0^2 p(t) \\ & + C_b d \sum_{t=0}^{t_0-1} \sum_{\tau=t(u/d-1)+1}^{\infty} \left[\tau - \frac{(u-d)t}{d} \right] g(\tau) p(t) \end{aligned} \tag{1}$$

and

$$\begin{aligned} T(t_0) = {} & \sum_{t=0}^{t_0-1} \sum_{\tau=1}^{t(u/d-1)} t \left(\frac{u}{d} \right) g(\tau) p(t) \\ & + \sum_{t=0}^{t_0-1} \sum_{r=t(u/d-1)+1}^{\infty} (t+\tau) g(\tau) p(t) \\ & + \sum_{t=t_0}^{\infty} t_0 \left(\frac{u}{d} \right) p(t), \end{aligned} \tag{2}$$

where $p(t)$ is the probability mass function of the discrete failure time distribution, $g(\tau)$ is the probability mass function of the discrete maintenance time distribution.

Giri considers that machine failures and repair times obey geometric distributions, i.e.

$$\begin{aligned} p(t) &= \begin{cases} 0, & t = 0 \\ (1-p)^{t-1} p, & t = 1, 2, 3, \ldots; 0 < p < 1 \end{cases} \\ g(\tau) &= \begin{cases} 0, & \tau = 0 \\ (1-q)^{\tau-1} q, & \tau = 1, 2, 3, \ldots; 0 < q < 1 \end{cases} \end{aligned} \tag{3}$$

Combining the above fault and repair time distributions with equations (1)-(2) gives the average total cost per cycle $U_I(t_0)$ and the average production-inventory cycle $T_I(t_0)$ to compute the average total cost per unit of time ETU_I, where t_0 denotes the machine production time. The formulas for $U_I(t_0), T_I(t_0)$, and ETU_I are presented as follows:

$$\begin{aligned} U_{\mathrm{I}}(t_0) = {} & C_p + C_f pq \sum_{t=1}^{t_0-1} (1-p)^{t-1} \sum_{\tau=1}^{\infty} \tau (1-q)^{\tau-1} \\ & + \frac{C_h u(u-d)p}{2d} \sum_{t=1}^{t_0-1} t^2 (1-p)^{t-1} \\ & + \frac{C_h u(u-d)}{2d} t_0^2 (1-p)^{t_0-1} \\ & + C_b dpq \left[\sum_{t=1}^{t_0-1} (1-p)^{t-1} \sum_{\tau=t(p/d-1)+1}^{\infty} \tau (1-q)^{\tau-1} \right. \\ & \left. - \frac{u-d}{d} \sum_{t=1}^{t_0-1} t (1-p)^{t-1} \sum_{\tau=t(p/d-1)+1}^{\infty} (1-q)^{\tau-1} \right], \end{aligned} \tag{4}$$

$$\begin{aligned} T_I(t_0) = {} & \frac{upq}{d} \sum_{t=1}^{t_0-1} t (1-p)^{t-1} \sum_{\tau=1}^{t(u/d-1)} (1-q)^{\tau-1} \\ & + pq \sum_{t=1}^{t_0-1} t (1-p)^{t-1} \sum_{\tau=t(p/d-1)+1}^{\infty} (1-q)^{\tau-1} \\ & + pq \sum_{t=1}^{t_0-1} (1-p)^{t-1} \sum_{\tau=t(p/d-1)+1}^{\infty} \tau (1-q)^{\tau-1} \\ & + \left(\frac{u}{d} \right) t_0 (1-p)^{t_0-1}, \end{aligned} \tag{5}$$

$$ETU_I = U_I(t_0)/T_I(t_0). \tag{6}$$

4.2 Steady-State Probability Model for Production-Inventory Systems (Model II)

In this section, we integrate machine states with inventory levels to construct a Markov model characterized by a state transition probability matrix $\boldsymbol{P}$, based on the following assumptions: The machine states are classified as 0 (complete failure), 1 (normal operation), and 2 (shutdown for rest). The corresponding inventory levels intervals for each machine state are $[0, 1, \cdots, I_{\max}]$, $[0, 1, \cdots, I_{\max} - 1]$, and $[1, 2, \cdots, I_{\max}]$, respectively. The system state is represented as $S = (M, I)$, encompassing $3I_{\max} + 1$ possible states. Consequently, two scenarios are excluded from consideration: full inventory during normal operation and zero inventory during shutdown for rest.

The structure of $\boldsymbol{P}$ is as follows:

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} & \boldsymbol{C} \\ \boldsymbol{D} & \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{G} & \boldsymbol{H} \end{pmatrix}, \tag{7}$$

where $\boldsymbol{P}$ is a square matrix of order $3I_{\max} + 1$, and the state transition rules for each submatrix are as follows:

$$\begin{aligned}
&\boldsymbol{A}: P\left\{(M=0,I) \to \left(M=0, I^{'}=\max\left(I-d,0\right)\right)\right\} = 1-q,\\
&\boldsymbol{B}: P\left\{(M=0,I) \to \left(M=1, I^{'}=\max(I-d,0)=0\right)\right\} = q,\\
&\mathbf{C}: P\left\{(M=0,I) \to \left(M=2, I^{'}=\max\left(I-d,0\right)>0\right)\right\} = q,\\
&\boldsymbol{D}: P\left\{(M=1,I) \to \left(M=0, I^{'}=\min\left(\max\left(I+u-d,0\right), I_{\max}\right)\right)\right\} = p,\\
&\boldsymbol{E}: P\left\{(M=1,I) \to \left(M=1, I^{'}=I+u-d<I_{\max}\right)\right\} = 1-p,\\
&\boldsymbol{F}: P\left\{(M=1,I) \to \left(M=2, I^{'}=I+u-d\geq I_{\max}\right)\right\} = 1-p,\\
&\mathbf{G}: P\left\{(M=2,I) \to \left(M=1, I^{'}=\max\left(I-d,0\right)=0\right)\right\} = 1\\
&\boldsymbol{H}: P\left\{(M=2,I) \to \left(M=2, I^{'}=\max\left(I-d,0\right)>0\right)\right\} = 1.
\end{aligned} \tag{8}$$

After obtaining the system state transition probability matrix, the steady-state probability $\boldsymbol{\pi}$ for each state is computed in accordance with equation (9):

$$\begin{cases} \pi P = \pi \\ \pi e = 1, \end{cases} \tag{9}$$

where $\boldsymbol{\pi} = \left(\pi_{(0,0)}, \cdots, \pi_{(0,I_{\max})}, \pi_{(1,0)} \cdots, \pi_{(1,I_{\max}-1)}, \pi_{(2,1)}, \cdots, \pi_{(2,I_{\max})}\right)$, and $\boldsymbol{e}$ is a $3I_{max} + 1$-dimensional column vector in which all elements are equal to 1.

Considering all states, the average total cost per unit of time ETU_{II} can be obtained as

$$\begin{aligned} ETU_{II} = &\sum_{I=0}^{I_{\max}} \pi_{(0,I)} \left(C_f + C_h I + C_b \delta\left(I=0\right)\right) \\ &+ \sum_{I=0}^{I_{\max}-1} \pi_{(1,I)} \left(C_p + C_h I\right) \\ &+ \sum_{I=1}^{I_{\max}} \pi_{(2,I)} \left(C_r + C_h I\right), \end{aligned} \tag{10}$$

where $\delta(I=0)$ is the indicator function.

The production cycle is defined as the time interval from when the inventory is zero and the machine is in production normal operation until the next return to this state.

The average production-inventory cycle T_{II} can be obtained as

$$T_{\Pi} = \frac{1}{\pi\left(1,0\right)}. \tag{11}$$

The average total cost per cycle U_{II} can be obtained as

$$U_{\Pi} = T_{\Pi} \times ETU_{\Pi}. \tag{12}$$

4.3 Reinforcement Learning-Based Decision Model for Production-Inventory Systems (Model III)

Reinforcement learning, as a vital subfield of machine learning, is fundamentally based on the concept of an agent that continuously interacts with its environment to acquire optimal decision-making policies. The theoretical foundation of this discipline is established upon the MDP, which serves as a systematic modeling framework for various decision-making challenges, particularly those necessitating dynamic decision-making processes such as inventory management.

Figure 2 illustrates a Reinforcement Learning approach based on MDP that employs the cost function $c(t)$ as a reward signal. This methodology effectively tackles decision-making challenges in production inventory control, such as identifying optimal production points and inventory levels, by continuously refining strategies to minimize long-term operational costs.

The fundamental components of the production-inventory research problem discussed in this section (state space, action space, inventory update rules, state transition rules, and cost function) are defined as follows:

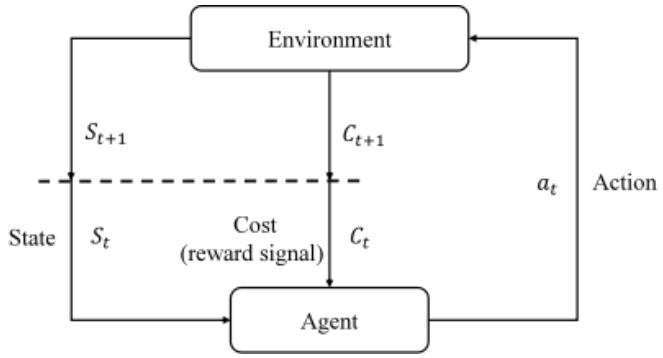

Figure 2. The interaction between agent and environment in an MDP.

4.3.1 State space

The machine state M and inventory level I are integrated into the system state $S = (M, I)$, where $I \in \{0, 1, 2, \cdots, I_{\max}\}$, $M \in \{0, 1, 2\}$, with 0, 1, and 2 representing complete failure, normal operation, and shutdown for rest, respectively. Similarly, scenarios involving full inventory with normal operation or zero inventory with shutdown for rest are excluded.

4.3.2 Action space

The action a ($a \in A$) that the machine can choose is to produce or not to produce, where $a = 1$ represents production and $a = 0$ represents no production. The machine is forced not to produce when in a complete failure state, while it can choose whether to produce as needed when in a normal operation or shutdown for rest state. The action space is as follows:

$$A = \begin{cases} \{0\} & M = 0 \\ \{0, 1\} & M = 1, 2, \end{cases} \tag{13}$$

where production during machine normal operation is referred to as loaded operation, and no production during machine normal operation is referred to as no-load operation.

4.3.3 Inventory update rule

For the inventory problem, when the machine action is set to production ($a = 1$), the inventory increment per time unit is $u - d$, but it does not exceed the maximum inventory level I_{max}. When the machine action is set to non-production ($a = 0$), the inventory decrement per time unit is d, but negative inventory levels are not allowed. Thus

$$I_{t+1} = \begin{cases} \min(I_t + u - d, I_{\max}) & a = 1 \\ \max(I_t - d, 0) & a = 0, \end{cases} \tag{14}$$

where I_t represents the inventory level at time t.

4.3.4 State transition rules

The state transition probability denotes the likelihood of the system transitioning from the current state to another state. The operational states of the machine are classified into three categories: complete failure, normal operation, and shutdown for rest. Each category is associated with distinct transition probabilities, wherein both failure and repair durations adhere to a geometric distribution. Assuming the initial machine state is normal operation ($M = 1$), the selected action is production ($a = 1$), and the inventory level is zero ($I = 0$), the state transition is illustrated in Figure 3.

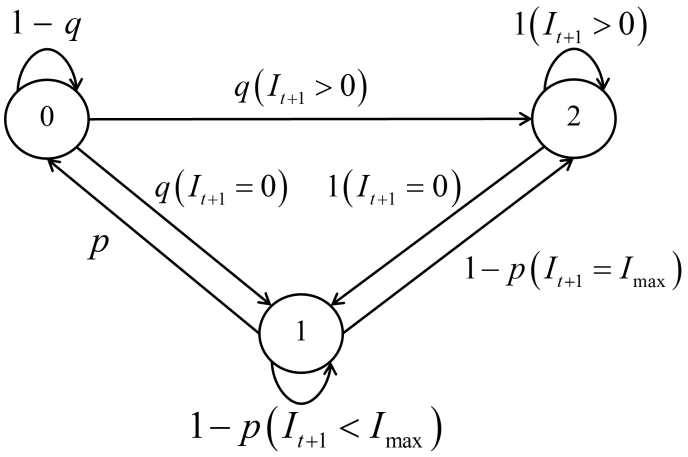

Figure 3. State transition diagram of the machine.

Let M_t denote the state of the machine at time t. As illustrated in Figure 3, when $M_t = 0$, the machine repair time follows a geometric distribution with parameter q. If the inventory level at the next time step is greater than 0 ($I_{t+1} > 0$), the machine transitions to the shutdown for rest state ($M_{t+1} = 2$) after repair. Conversely, if the inventory level at the next time step is 0 ($I_{t+1} = 0$), the machine transitions to the normal operation state ($M_{t+1} = 1$) after repair. The specific transition probabilities are as follows:

$$\begin{cases} P(M_{t+1} = 2 \mid M_t = 0) = q, & I_{t+1} > 0, \\ P(M_{t+1} = 1 \mid M_t = 0) = q, & I_{t+1} = 0, \\ P(M_{t+1} = 0 \mid M_t = 0) = 1 - q. \end{cases} \tag{15}$$

When $M_t = 1$, the machine failure time follows a geometric distribution with parameter p. When the inventory reaches its maximum level at the next time step ($I_{t+1} = I_{max}$), the machine enters the shutdown

for rest state ($M_{t+1} = 2$). The specific transition probabilities are as follows:

$$\begin{cases} P\left(M_{t+1}=1|M_t=1\right)=1-p, & I_{t+1}<I_{\max}, \\ P\left(M_{t+1}=2|M_t=1\right)=1-p, & I_{t+1}=I_{\max}, \\ P\left(M_{t+1}=0|M_t=1\right)=p. \end{cases} \tag{16}$$

When $M_t = 2$, it is stipulated that the machine cannot transition to a complete failure state at the next moment. If the inventory level at the next moment is zero ($I_{t+1} = 0$), the machine enters the normal operation state ($M_{t+1} = 1$). The specific transition probabilities are as follows:

$$\begin{cases} P\left(M_{t+1}=1|M_t=2\right)=1, & I_{t+1}=0, \\ P\left(M_{t+1}=2|M_t=2\right)=1, & I_{t+1}>0. \end{cases} \tag{17}$$

4.3.5 Cost function

The cost function established in this section includes the single production cost (C_p), inventory holding cost per unit of product per unit of time (C_h), single stockout penalty cost (C_b), maintenance cost of machine per time unit (C_f), and maintenance cost during machine shutdown for rest (C_r). At time t, the system generates the instantaneous cost $C(S_t, a_t)$ based on the current system state S_t and the selected action $a(t)$. The specific expression is as follows:

$$\begin{aligned} C\left(S_t,a_t\right) = {} & C_p a_t + C_h I_t + C_b \delta\left(I_t=0, M_t=0\right) \\ & + C_f \delta\left(M_t=0\right) + C_r \delta\left(M_t=2\right), \end{aligned} \tag{18}$$

where $S_t = \left(M_t, I_t\right)$ represents the system state at time t, M_t denotes the machine state at time t, I_t indicates the inventory level at time $t, a_t(a_t \in A)$ stands for the action selected by the machine at time t, and $\delta(\cdot)$ is the indicator function.

The definitions of the average production-inventory cycle (T_{III}), the average total cost per cycle (U_{III}), and the average total cost per unit of time (ETU_{III}) for Model III are presented below.

The average production-inventory cycle (T_{III}) represents the mean time interval between two consecutive production start times ($I = 0, M = 1, a = 1$) during long-term system operation. The formula is given as

$$T_{III} = \frac{1}{N}\sum_{k=1}^{N} T_k, \tag{19}$$

where N denotes the total number of observations within the complete production-inventory cycle of Model III, and T_k represents the time interval corresponding to the k-th cycle.

The average total cost per cycle (U_{III}) represents the average total cost per production-inventory cycle under long-term system operation. The formula is given as

$$U_{III} = \frac{1}{N}\sum_{k=1}^{N} U_k, \tag{20}$$

where U_k denotes the total cost associated with the k-th cycle in Model III.

ETU_{III} represents the average cost per time step during long-term system operation, which is the ratio of the average total cost per cycle U_{III} to the average production-inventory cycle T_{III}. The formula is given as

$$ETU_{III} = \frac{U_{III}}{T_{III}} \tag{21}$$

In this section, two distinct inventory production schemes are modeled to determine the average total cost per unit of time.

Case 1: The machine functions under load during its normal operation state; however, production is suspended in instances of complete failure, shutdown for rest, or when inventory levels attain their maximum capacity.

Case 2: The system is capable of selecting production actions based on inventory levels during both normal operation and shutdown for rest states of the machine. However, it refrains from producing when the machine experiences a complete failure or when the inventory level reaches its maximum capacity.

4.4 Q-learning Algorithm

This section approximates the optimal decision-making policy by updating the state-action value function $Q(S_t, a_t)$ through the Q-learning algorithm, a model-free Reinforcement Learning method grounded in value iteration. By facilitating real-time interactions between the agent and its environment, this approach effectively addresses the

Bellman optimality equation without necessitating prior knowledge of the system dynamics. The algorithm employs an ε-greedy policy ($\varepsilon \leftarrow \varepsilon \cdot 0.999$) to strike a balance between exploration and exploitation; it initially broadens the decision space via high-probability random exploration while progressively enhancing the focus on optimal actions as learning advances.

In production-inventory problems, the algorithm assesses the expected operational cost associated with each decision point as a criterion for the action-value function. At time t, after the agent executes action a_t in state S_t, it observes the immediate cost $C(S_t, a_t)$ and transitions to the next state S_{t+1}. Subsequently, it iteratively updates the Q-value in accordance with the Bellman equation, as follows:

$$Q(S_t, a_t) \leftarrow Q(S_t, a_t) + \alpha \left[C(S_t, a_t) + \gamma \min_{a \in A} Q(S_{t+1}, a) - Q(S_t, a_t) \right], \quad (22)$$

where the learning rate α governs the magnitude of parameter adjustments, and the discount factor γ regulates the weight assigned to future costs. This iterative process ultimately converges to the optimal action-value function, thereby providing a quantitative foundation for production decision-making. Meanwhile, during the calculation of the minimum average total cost per unit of time, the Bayesian optimization algorithm is employed for hyperparameter selection, with the search ranges of $(\alpha, \gamma, \varepsilon)$ defined as $[0, 1]$. The specific algorithm is as shown in Algorithm 1.

Algorithm 1: Q-learning algorithm for production inventory system

Data: $n, p, q, u, d, C_p, C_h, C_b, C_f, C_r$, episodes, steps
Result: T, U, ETU
for $i = 1$ *to length*(n) **do**
  $I_{\max} \leftarrow n(i)$;
  Hyperparameter selection using Bayesian Optimization;
  Optimal params ← Use the bayesopt function;
  Training with optimal parameters and updating function with Q table;
  for $ep = 1$ *to episodes* **do**
    Set initial machine state M and inventory level I, compute current state;
    for $t = 1$ *to steps* **do**
      Choose actions according to different case combined with ε-greedy policy;
      Compute the next state according to the state transition rule;
      Compute the immediate cost C;
      Compute cumulative costs over a cycle: total cost ← total cost + C;
      Detect and record cycle duration and cycle cost;
      Iterative updating of $Q(S_t, a_t)$ value combined with Bellman equation;
      Update the state: $M \leftarrow M_{next}$, $I \leftarrow I_{next}$, current state ← next state;
    end
    $\varepsilon \leftarrow \varepsilon \cdot 0.999$;
  end
  $T \leftarrow$ mean of cycle duration, $U \leftarrow$ mean of cycle cost, $ETU \leftarrow U/T$;
end
return T, U, ETU

5 Numerical Results

5.1 Comparison of Results Among Models When Maintenance Cost During Machine Shutdown for Rest are Zero ($C_r = 0$)

In Model I, since maintenance cost during machine shutdown for rest are not involved, the model parameters are set as follows: $p = 0.4$, $q = 0.8$, $u = 30$, $d = 5$, $C_p = 20$, $C_h = 5$, $C_b = 100$, $C_f = 50$. Figures 4, 5 and 6 illustrate the variations in the average total cost per unit of time, the average production-inventory cycle, and the average total cost per cycle with respect to inventory levels.

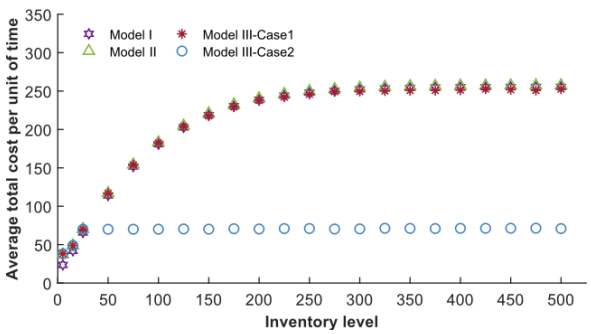

Figure 4. Average total cost per unit of time under different models ($C_r = 0$).

As illustrated in Figures 4, 5 and 6, Model III-Case2 (where the machine has the flexibility to decide whether to produce during both normal operation and shutdown for rest states) demonstrates notable differences in terms of average total cost per time

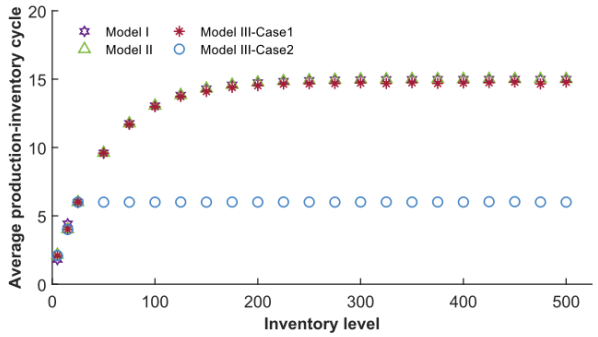

Figure 5. Average production-inventory cycle under different models ($C_r = 0$).

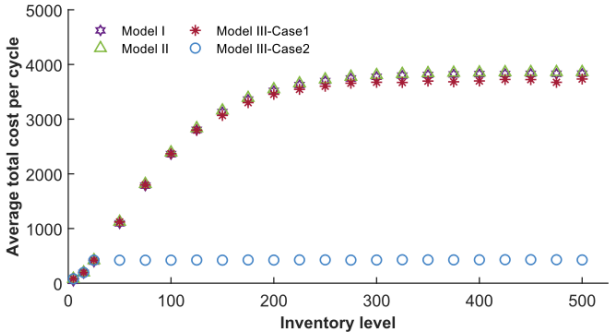

Figure 6. Average total cost per cycle under different models ($C_r = 0$).

unit, average production-inventory cycle, and average total cost per cycle when compared to the other three models. At low inventory levels ($I \leq 25$), the outcomes across all four models are relatively similar. However, as inventory levels rise, Model III-Case2 exhibits a more rapid convergence in ETU, T, and U metrics alongside reduced costs. In contrast, the remaining three models display a continuous upward trend in these three metrics at medium inventory levels ($25 < I \leq 250$), which eventually stabilize at high inventory levels ($I > 250$) with numerically comparable results. This suggests that the Reinforcement Learning approach facilitates dynamic production adjustments based on cost considerations, thereby achieving a lower long-run average total cost per time unit more effectively across varying inventory levels. When I_{max} is 50, 100, 200, 300, 400 and 500, the convergence plots of the average total cost per unit of time for Model III-Case1 and Case2 are shown in Figures A1 and A2 of Appendix A.1. The specific numerical results for ETU and T for each model are detailed in Table 1.

As illustrated in Table 1, Model III-Case2 exhibits significant advantages at medium to high inventory levels. Its ETU value is consistently maintained within the range of 70-72, while the T value remains stable at 6. This stability indicates the model's capability to adjust production strategies in response to varying inventory levels, thereby stabilizing production cycles and minimizing the average total cost per time unit. In contrast, the other three models demonstrate noticeable upward trends at medium inventory levels with slower convergence rates. This further underscores that allowing machines to alternate between normal operation and shutdown for rest states can effectively reduce the average total cost per time unit during production.

5.2 Comparison of Model Outcomes Considering Non-Zero Maintenance Cost During Machine Shutdown for Rest ($C_r \neq 0$)

The introduction of maintenance cost during machine shutdown for rest, which accounts for the ongoing maintenance expenses incurred when machines are in a shutdown for rest state, better aligns with real-world scenarios. The parameter values are set as: $p = 0.4$, $q = 0.8$, $u = 30$, $d = 5$, $C_p = 20$, $C_h = 5$, $C_b = 100$, $C_f = 50$, $C_r = 10$. Since Model I does not involve maintenance cost during machine shutdown for rest, this section only compares the numerical results of Model II, Model III-Case1, and Model III-Case2. The results of various indicators under different inventory levels are illustrated in Figures 7, 8 and 9.

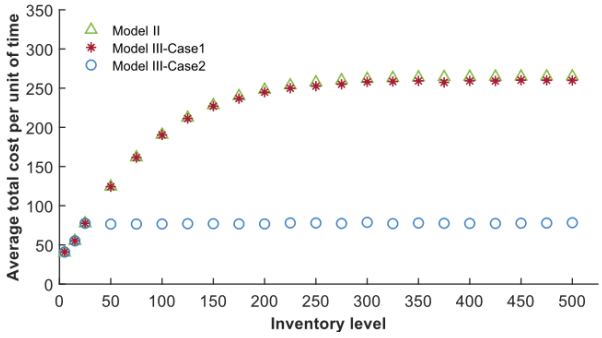

Figure 7. Average total cost per unit of time under different models ($C_r \neq 0$).

As illustrated in Figures 7, 8 and 9, Model III-Case2 continues to exhibit significant advantages under the condition of $C_r \neq 0$. At medium to high inventory levels ($I \geq 75$), the model can be adjusted according to strategic considerations, facilitating cost convergence with stable cycles. In comparison to Model II and Model III-Case1, it demonstrates a faster convergence rate and lower associated costs. The performance of the other two models is comparable; all indicators show a gradual increase followed by stabilization as inventory levels rise, with stabilization occurring only at elevated inventory levels ($I > 300$). When I_{max} is 50, 100, 200,

Table 1. Numerical results ($C_r = 0$).

I	ETU_I	ETU_{II}	$ETU_{III-Case1}$	$ETU_{III-Case2}$	T_I	T_{II}	$T_{III-Case1}$	$T_{III-Case2}$
5	23.58	37.11	38.02	38.06	1.81	2.15	2.10	2.10
15	42.03	48.76	48.82	48.88	4.42	4.01	4.00	4.00
25	65.83	70.00	70.06	70.08	6.00	6.00	6.00	6.00
50	114.06	116.87	116.80	70.04	9.60	9.60	9.58	6.00
75	151.79	154.18	153.72	70.05	11.76	11.76	11.69	6.00
100	180.84	183.05	182.38	70.14	13.06	13.06	12.99	6.00
125	202.68	204.80	203.77	70.42	13.83	13.83	13.73	6.01
150	218.73	220.79	217.69	70.20	14.30	14.30	14.09	6.00
175	230.26	232.30	229.39	70.72	14.58	14.58	14.40	6.01
200	238.39	240.42	237.87	70.39	14.75	14.75	14.52	6.01
225	244.03	246.05	242.11	70.93	14.85	14.85	14.64	6.01
250	247.89	249.89	245.72	71.00	14.91	14.91	14.66	6.01
275	250.49	252.50	249.18	70.36	14.95	14.95	14.68	6.00
300	252.23	254.23	249.57	70.54	14.97	14.97	14.72	6.01
325	253.37	255.37	250.16	71.30	14.98	14.98	14.67	6.02
350	254.13	256.13	250.77	71.14	14.99	14.99	14.74	6.01
375	254.62	256.62	250.78	70.93	14.99	15.00	14.69	6.01
400	254.94	256.93	251.22	71.23	15.00	15.00	14.72	6.01
425	255.14	257.15	252.73	71.27	15.00	15.00	14.74	6.03
450	255.27	257.28	251.83	71.34	15.00	15.00	14.79	6.04
475	255.36	257.36	250.77	71.43	15.00	15.00	14.65	6.01
500	255.41	257.41	252.67	70.92	15.00	15.00	14.78	6.01

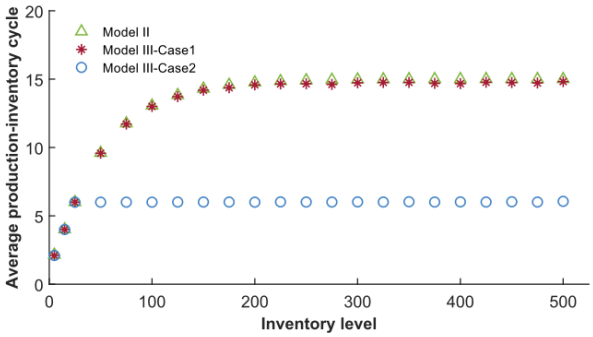

Figure 8. Average production-inventory cycle under different models ($C_r \neq 0$).

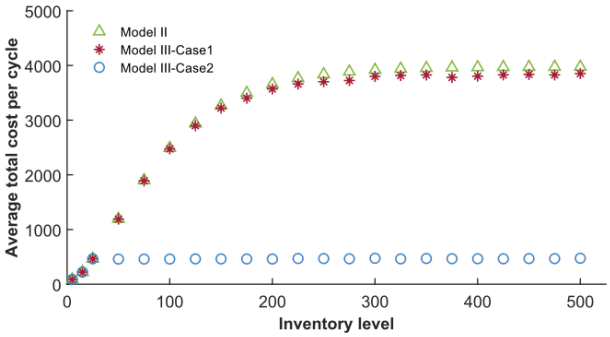

Figure 9. Average total cost per cycle under different models ($C_r \neq 0$).

300, 400 and 500, the convergence plots of the average total cost per unit of time for Model III-Case1 and Case2 are presented in Figures A3 and A4 of Appendix A.2. The specific numerical results are presented in Table 2.

As illustrated in Table 2, the average total cost per unit of time of Model III-Case2 remains stable within the range of 76-80, while the average production-inventory cycle consistently maintains a duration of 6. Both metrics are lower than their corresponding indicators for Model II and Model III-Case1. This demonstrates that under the condition of $C_r \neq 0$, Model III-Case2 continues to demonstrate significant advantages even in high inventory environments.

Based on the Q-values derived from selecting various actions in the Q-learning algorithm, Figures 10 and 11 illustrate the optimal production strategies identified for $I_{max} = 50$ and $I_{max} = 100$, respectively. As depicted in Figure 10, when $I_{max} = 50$ is considered, the machine adopts an intermittent production strategy (with a production interval of 4) during shutdown for rest states characterized by low inventory levels ($I \leq 20$). Conversely, during normal operations, the machine refrains from producing at low inventory levels ($I \leq 20$) but opts

Table 2. Numerical results ($C_r > 0$).

I	ETU_{II}	$ETU_{III-Case1}$	$ETU_{III-Case2}$	T_{II}	$T_{III-Case1}$	$T_{III-Case2}$
5	40.11	40.83	40.94	2.15	2.10	2.10
15	55.03	55.10	55.11	4.01	4.00	4.00
25	77.50	77.56	77.54	6.00	6.00	6.00
50	124.37	124.28	76.57	9.60	9.58	6.00
75	161.68	161.42	76.61	11.76	11.71	6.00
100	190.55	190.33	76.71	13.06	13.00	6.00
125	212.30	211.28	77.04	13.83	13.72	6.01
150	228.29	226.95	77.01	14.30	14.19	6.01
175	239.80	236.71	76.76	14.58	14.37	6.01
200	247.92	244.87	76.76	14.75	14.58	6.00
225	253.55	249.88	78.00	14.85	14.65	6.02
250	257.39	252.55	77.96	14.91	14.66	6.01
275	260.00	254.94	77.35	14.95	14.61	6.01
300	261.73	258.02	78.76	14.97	14.73	6.02
325	262.87	258.52	77.14	14.98	14.76	6.01
350	263.63	259.25	77.92	14.99	14.76	6.03
375	264.12	257.37	77.54	15.00	14.69	6.01
400	264.43	259.26	77.39	15.00	14.67	6.02
425	264.65	259.18	77.33	15.00	14.77	6.01
450	264.78	260.15	77.69	15.00	14.75	6.03
475	264.86	260.00	77.85	15.00	14.72	6.01
500	264.91	259.87	78.33	15.00	14.82	6.06

to produce at higher inventory levels ($I = 40$ and 45). Figure 11 further illustrates that when $I_{max} = 100$ is applicable, the machine also selects intermittent production (with a production interval of 4) while in shutdown for rest state with low to medium inventory levels ($I \leq 60$). During normal operational phases, the production points exhibit greater dispersion; however, they predominantly cluster around medium inventory levels ($40 \leq I \leq 55$).

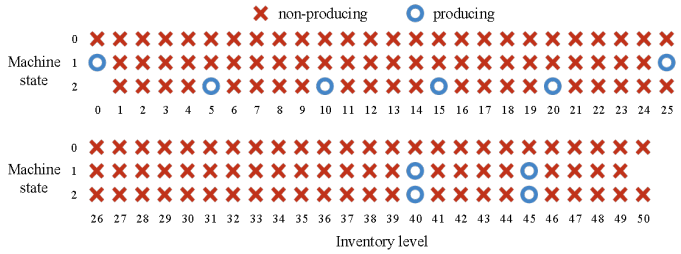

Figure 10. Optimal production strategy at $I_{max} = 50$.

5.3 Sensitivity of model results for different types of costs

Based on Section 5.2 parameters, we set the inventory level at $I = 300$. Take the values of C_p, C_h, C_b, C_f, C_r within 0-100 respectively. Meanwhile, the remaining parameters are set to the default values in Table 3.

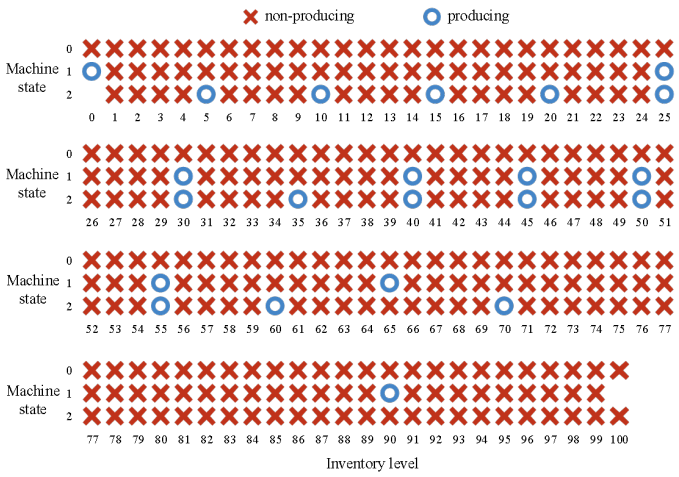

Figure 11. Optimal production strategy at $I_{max} = 100$.

Table 3. Parameter value.

p	q	u	d	C_p	C_h	C_b	C_f	C_r
0.4	0.8	30	5	20	5	100	50	10

This allowed us to calculate the average total cost per unit of time across different cost scenarios for each model. The results are presented in Figure 12 (where (a) is Model II, (b) is Model III-Case1, and (c) is Model III-Case2).

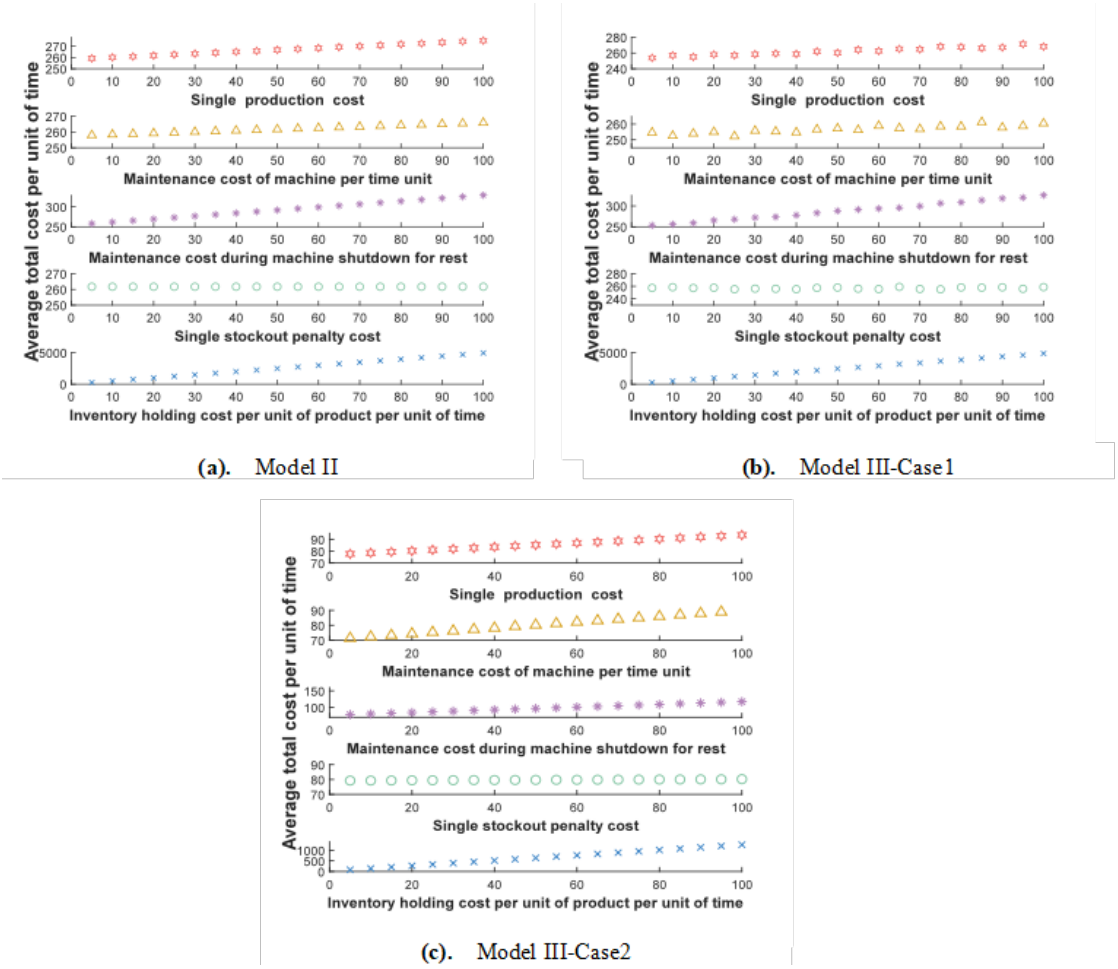

Figure 12. Average total cost per unit of time for different models under various cost variations.

Figure 12 shows that the average total cost per unit of time increases with higher C_p and C_f values, rises linearly with increasing C_r and C_h values, but remains stable across different C_b values. Notably, Model III-Case2 consistently achieves the lowest costs among all models, demonstrating both cost stability against parameter variations and effective cost optimization through production strategy adjustments.

The results further substantiate the robustness of the previous conclusions: regardless of variations in unit costs, Model III-Case2 consistently remains optimal, indicating that its optimization capability is not merely coincidental. Overall, ETU increases as most individual costs rise, which aligns with the linear relationship posited in the cost function. From a sensitivity analysis perspective, total cost exhibits the greatest sensitivity to changes in C_h and comparatively less sensitivity to fluctuations in C_b. The latter does not induce a significant alteration in total cost, it is plausible that equipment has been repaired prior to experiencing shortages and resumes operation at the onset of the subsequent production cycle, thereby mitigating high penalty costs.

6 Conclusion

This study develops a discrete-time production-inventory decision model and innovatively incorporates RL algorithms into the optimization of traditional EMQ model production strategies. This approach highlights the advantages of intelligent decision-making methods in dynamic production environments. The findings indicate that the Q-learning-based Model III-Case2 exhibits significant benefits across both cost scenarios, enabling the average total cost per unit of time to converge to a fixed interval at an accelerated rate while maintaining a consistent average production-inventory cycle. In comparison to the traditional EMQ model, it achieves a cost reduction of approximately 30–70%. When the inventory levels exceed $u - d$ units, the Reinforcement Learning model can flexibly adjust production decisions, overcoming the shortcomings of traditional models that rely on fixed strategies, which often result in slower cost convergence and

higher costs. This approach successfully balances the relationship between the average total cost per unit of time and production strategies, thereby facilitating optimized dynamic production control with respect to costs.

In practical management applications, managers can integrate the model into production management software systems, and dynamically generate optimal production decisions by collecting real-time machine status, inventory levels and market demand data. The adaptability of the model enables managers to effectively deal with uncertain factors such as sudden machine failures, thus helping enterprises to reduce operating costs and improve equipment utilization. However, there are still some limitations in this study. The model assumes that machine failure and repair time obey geometric distribution and is a single-machine single-product production scenario, which may be different from the complex failure mode in actual production. Future studies can consider multi-machine multi-product production scenarios and introduce assumptions about general distributions of machine failure times, or consider more complex production-inventory-repair policy models.

Appendix

A.1 Convergence Plots When Maintenance Cost During Machine Shutdown for Rest are Zero ($C_r = 0$)

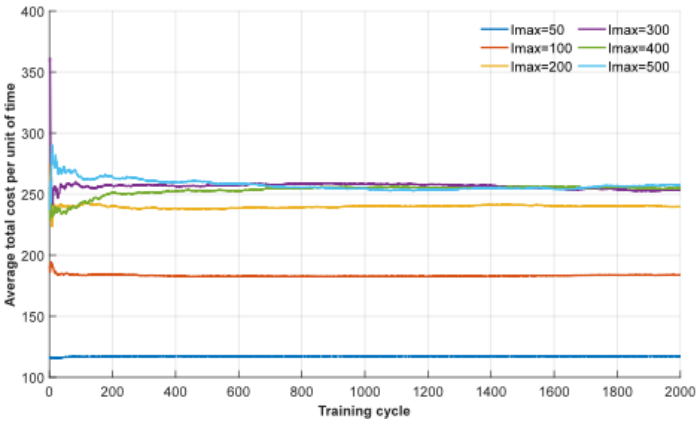

Figure A1. Convergence plots of average total cost per unit of time under Model III-Case1 ($C_r = 0$).

A.2 Convergence Plots Considering Non-Zero Maintenance Cost During Machine Shutdown for Rest ($C_r \neq 0$)

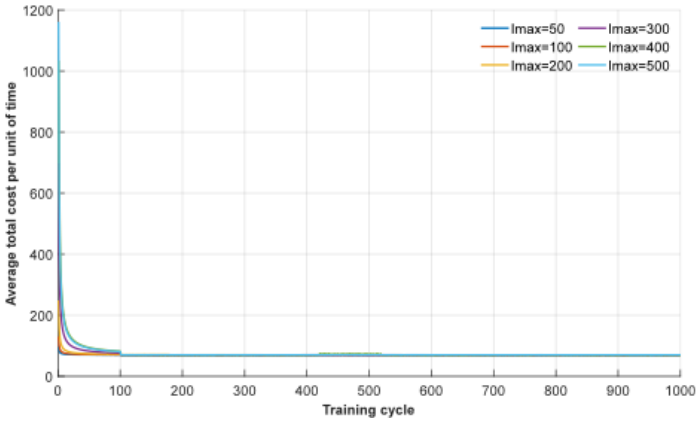

Figure A2. Convergence plots of average total cost per unit of time under Model III-Case2 ($C_r = 0$).

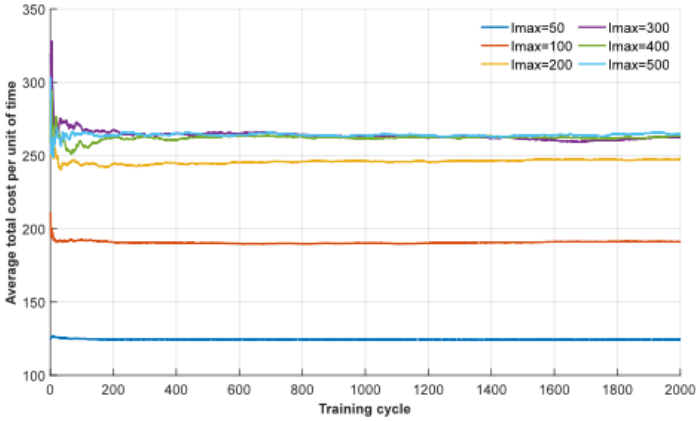

Figure A3. Convergence plots of average total cost per unit of time under Model III-Case1($C_r \neq 0$).

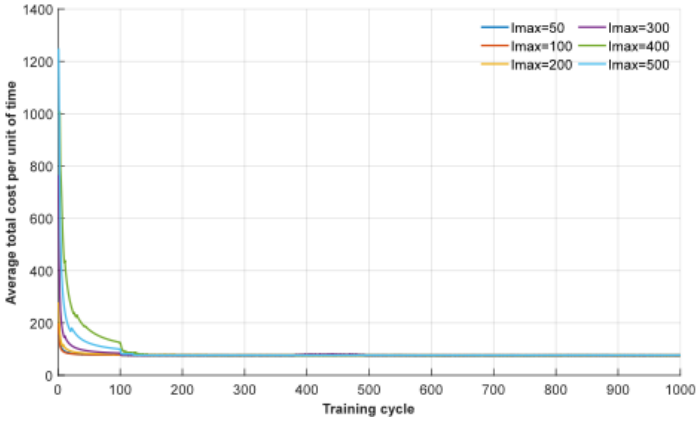

Figure A4. Convergence plots of average total cost per unit of time under Model III-Case2($C_r \neq 0$).

Data Availability Statement

Data will be made available on request.

Funding

This work was supported in part by the Shijiazhuang Science and Technology Project under Grant 241790737A; in part by the Natural Science Foundation of Hebei Province under Grant G2025203034; in part

by the Shanxi Provincial Basic Research Program Youth Project under Grant 202403021212004.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Harris, F. W. (1990). How many parts to make at once. *Operations research, 38*(6), 947-950. [Crossref]

[2] Jaber, M. Y., & Bonney, M. (1999). The economic manufacture/order quantity (EMQ/EOQ) and the learning curve: past, present, and future. *International journal of production economics, 59*(1-3), 93-102. [Crossref]

[3] Giri, B. C., & Dohi, T. (2006). Discrete-time economic manufacturing quantity model with stochastic machine breakdown and repair. In *Reliability Modeling, Analysis And Optimization* (pp. 81-106). [Crossref]

[4] Giri, B. C., Yun, W. Y., & Dohi, T. (2005). Optimal design of unreliable production–inventory systems with variable production rate. *European Journal of Operational Research, 162*(2), 372-386. [Crossref]

[5] Chiu, Y. S. P., Liu, S. C., Chiu, C. L., & Chang, H. H. (2011). Mathematical modeling for determining the replenishment policy for EMQ model with rework and multiple shipments. *Mathematical and Computer Modelling, 54*(9-10), 2165-2174. [Crossref]

[6] Chiu, K. C., Yeh, C. W., & Fang, C. C. (2010, December). An EMQ model with time-varying demand over the product life cycle. In *2010 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 1683-1687). IEEE. [Crossref]

[7] Sarkar, B., Mandal, P., & Sarkar, S. (2014). An EMQ model with price and time dependent demand under the effect of reliability and inflation. *Applied Mathematics and Computation, 231*, 414-421. [Crossref]

[8] Borrero, J. S., & Akhavan-Tabatabaei, R. (2013). Time and inventory dependent optimal maintenance policies for single machine workstations: An MDP approach. *European Journal of Operational Research, 228*(3), 545-555. [Crossref]

[9] Zhang, N., Cai, K. Q., Deng, Y. J., & 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 and System Safety, 241*, 109694. [Crossref]

[10] Zhang, N., Tian, S., Xu, J., Deng, Y., & Cai, K. (2023). Optimal production lot-sizing and condition-based maintenance policy considering imperfect manufacturing process and inspection errors. *Computers & Industrial Engineering, 177*, 108929. [Crossref]

[11] Han, R., Ma, X., Yang, L., Cao, H., Guo, H., & Lu, H. (2024, July). Integrated optimization model of economic manufacturing quantity and hybrid condition-based maintenance for continuous-production systems. In *IET Conference Proceedings CP886* (Vol. 2024, No. 12, pp. 1248-1254). Stevenage, UK: The Institution of Engineering and Technology. [Crossref]

[12] Tan, B., Karabağ, O., & Khayyati, S. (2023). Production and energy mode control of a production-inventory system. *European Journal of Operational Research, 308*(3), 1176-1187. [Crossref]

[13] Pazouki, M., Jaber, M. Y., & Afshari, H. (2025). Linking forward and backward product quality in a manufacturing/remanufacturing inventory system with price-quality-dependent demand and return rates. *Computers and Industrial Engineering, 204*, 111072. [Crossref]

[14] Li, J., Hu, L., & Zhou, Y. (2025). Reliability design and inventory optimization for production-inventory systems considering market demand satisfaction ability. *International Journal of General Systems*, 1-28. [Crossref]

[15] Wu, G., de Carvalho Servia, M. Á., & Mowbray, M. (2023). Distributional reinforcement learning for inventory management in multi-echelon supply chains. *Digital Chemical Engineering, 6*, 100073. [Crossref]

[16] Hubert, S., Meintschel, J., Bleidorn, D., Ortmanns, Y., & Wallrath, R. (2023). Production scheduling using deep reinforcement learning and discrete event simulation. *Chemie Ingenieur Technik, 95*(7), 1003-1011. [Crossref]

[17] Zhou, Y., Guo, K., Yu, C., & Zhang, Z. (2024). Optimization of multi-echelon spare parts inventory systems using multi-agent deep reinforcement learning. *Applied Mathematical Modelling, 125*, 827-844. [Crossref]

[18] Tian, R., Lu, M., Wang, H., Wang, B., & Tang, Q. (2024). IACPPO: A deep reinforcement learning-based model for warehouse inventory replenishment. *Computers & Industrial Engineering, 187*, 109829. [Crossref]

Renfang Wang is a postgraduate student at the School of Science at Yanshan University in China. His research interests include stochastic system reliability modeling and performance evaluation. (Email: wrf011110@163.com)

Yufei Gong received the Master degree in probability and statistics from the Beijing Jiaotong University, Beijing, China, in 2020, and the Ph.D. degree in system safety and reliability from the Troyes University of Technology, Troyes, France, in 2024. She is currently a Lecturer with the School of Science, Yanshan University. Her research interests include remaining useful life prognosis, predictive maintenance. (Email: yufeigong@ysu.edu.cn)

Peng Su received the Ph.D. degree in mathematics from the School of Mathematics, Southeast University, Nanjing, China, in 2020. He is currently an Associate Professor with the School of Economics and Management, North University of China. His research interests include stochastic modeling, Fuzzy reliability modeling, V2G technology research. (Email: supeng@nuc.edu.cn)

Linmin Hu received the Ph.D. degree in operations research and management science from the School of Economics and Management, Yanshan University, China, in 2014, where he is currently a professor with the Department of Applied Mathematics. He has published more than 50 papers in journals, including Applied Mathematical Modeling, Computers and Industrial Engineering, Reliability Engineering and System Safety and other journals. His research interests include system reliability, operations research and stochastic models. (Email: linminhu@ysu.edu.cn)

Xin Jiang is a PhD candidate at the School of Economics and Management at Yanshan University. Her research interests include reliability mathematics, reliability analysis of complex systems, and the joint optimization models of maintenance and spare parts inventory. (Email: jjjxin0919@163.com)