

PRODUCTION–INVENTORY COORDINATION UNDER UNCERTAIN DEMAND

Li, W.[#]

School of Business, Ewha Womans University, Seoul 03760, Republic of Korea

E-Mail: liwei@ewhain.net ([#] Corresponding author)

Abstract

Coordinating production and inventory decisions under uncertain demand remains difficult, as most existing approaches rely on predefined demand assumptions and loosely connect simulation with optimization. This study develops a reinforcement learning framework to support joint decision-making in such environments. The system is described as a Markov decision process, where demand dynamics are incorporated into the state representation and linked to a profit-oriented reward structure. Policy learning is implemented using a modified proximal policy optimization scheme to improve stability and responsiveness under demand fluctuations. The learning process is embedded within a discrete-event simulation environment, allowing continuous interaction between decision updates and system evolution. Experiments across multiple demand scenarios show that the proposed approach achieves consistently higher profit levels while reducing stockout risk and improving inventory stability compared with conventional policies and standard reinforcement learning methods.

(Received in January 2026, accepted in April 2026. This paper was with the author 1 month for 2 revisions.)

Key Words: Uncertain Demand, Production–Inventory–Profit Coordination, Deep Reinforcement Learning, Discrete-Event Simulation, Adaptive Optimization, Production System Simulation

1. INTRODUCTION

The intensification of global supply chain fluctuations [1, 2] and the trend of market demand fragmentation [3] make the coordinated optimization of production planning, inventory control, and profit maximization a key factor for manufacturing enterprises to enhance core competitiveness [4]. Market demand often exhibits characteristics of trend, seasonality, and sudden fluctuations. Traditional production simulation and optimization methods are difficult to adapt to such dynamic uncertainty [5, 6], resulting in imbalance between production and inventory decisions, and constraining the improvement of enterprise profit and operational stability. Existing studies still have obvious gaps. Traditional coordinated modelling mostly adopts a hierarchical optimization mode [7], in which production planning is formulated first, then inventory is regulated, and finally profit is calculated. This approach cannot achieve dynamic coordination among the three and is prone to generating suboptimal solutions and difficult to adapt to dynamic changes in demand distribution. Existing achievements of deep reinforcement learning in production simulation [8, 9] mostly ignore the accurate capture of demand temporal features, and the coupling degree between policy training and the simulation environment is insufficient, resulting in low training efficiency and lack of robustness. The integration of production simulation and intelligent optimization is mostly in an offline separated state [10, 11], lacking a real-time interactive integrated mechanism, which cannot realize the online adaptive update of strategies and is difficult to cope with sudden changes in market demand. These problems together constitute the core bottleneck of coordinated optimization of production–inventory–profit under uncertain demand, and it is urgent to construct a new simulation optimization paradigm to solve them.

To solve the above problems, this paper constructs a production–inventory–profit coordinated simulation optimization framework adapted to uncertain demand, realizes the dynamic coordination of production planning, inventory control, and profit maximization,

improves the profit level and demand fluctuation resistance of the production system, provides a new intelligent optimization paradigm for the field of production simulation, and provides technical support for dynamic decision-making of manufacturing enterprises. The main contributions are as follows:

(1) A refined Markov Decision Process (MDP) modelling method for production–inventory–profit coordination is proposed. It breaks through the limitations of traditional hierarchical modelling, introduces a Long Short-Term Memory (LSTM) network to encode demand temporal features, and designs a dynamically profit-oriented reward function, effectively solving the problems of incomplete state representation and coarse reward design in traditional models, and realizing the precise matching between state features and profit optimization objectives.

(2) An improved adaptive deep reinforcement learning algorithm is designed. Aiming at the continuous production quantity action space, the network structure and training strategy are optimized, and attention mechanism, prioritized experience replay, and adaptive learning rate adjustment strategies are incorporated, significantly improving the adaptability of the policy to demand mutation and training stability, and solving the problems of low training efficiency and insufficient robustness of traditional algorithms.

(3) A simulation–reinforcement learning integrated coupling mechanism is constructed. The interaction interface between the discrete-event simulation environment and the deep reinforcement learning algorithm is optimized, realizing the real-time linkage between policy training and simulation verification, breaking the limitations of the traditional offline separated integration mode, and improving optimization efficiency and policy adaptability.

The remainder of this paper is organized as follows. Section 2 elaborates the refined MDP modelling method of the production–inventory–profit coordination system; Section 3 proposes the specific design scheme of the improved adaptive deep reinforcement learning algorithm; Section 4 constructs the simulation–reinforcement learning integrated coupling environment and explains the design details of each core simulation module; Section 5 verifies the effectiveness and superiority of the proposed method through multi-scenario simulation experiments and sensitivity analysis; Section 6 summarizes the main conclusions and research contributions of this paper, and provides reference for subsequent related research.

2. MDP MODELLING OF THE PRODUCTION–INVENTORY SYSTEM

The research object of this paper is a single-stage production–inventory system. The system consists of a production unit, a finished goods warehouse, and an external demand side. To ensure modelling rationality and simulation realism, the core assumptions are set as follows. The production process has a fixed lead time, and the production quantity is constrained by both minimum production batch size and maximum production capacity. When inventory is insufficient, backordering is adopted, and shortage cost is incurred. Production cost consists of fixed cost and unit variable cost, and inventory holding cost is calculated based on the end-of-period inventory level. Market demand is a non-stationary stochastic process with characteristics of trend, seasonality, and sudden fluctuations. The above assumptions define the system operation boundary and dynamic characteristics, providing basic constraints for the subsequent MDP modelling of coordinated optimization.

Fig. 1 shows the refined MDP interaction framework of the production–inventory–profit coordination system. The innovative MDP modelling of the system breaks through the traditional design paradigm from two aspects: state space and action space. The state space takes current inventory level, work-in-process quantity, backlog of delayed orders, periodic temporal features, and cumulative profit as basic variables. The core innovation lies in

introducing an LSTM encoder to extract temporal features from the historical demand sequence of the past k periods and integrating the encoded features with the basic state to form a high-dimensional dynamic state vector. The state fusion and normalization processes follow:

$$s_t = W_h h_t + W_s s_t^{base} + b \quad (1)$$

$$\hat{x}_t = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (2)$$

where h_t is the demand temporal feature output by LSTM, s_t^{base} is the basic state vector, W_h and W_s are fusion weights, b is the bias term, and \hat{x}_t is the normalized state component. The action space is the continuous production quantity in each period. A constraint mapping layer is used to project the output of the policy network to the feasible interval. The mapping rule is: $a_t = a_{min} + \sigma(a_t^{net})(a_{max} - a_{min})$. At the same time, a feasibility penalty is imposed on the original action that exceeds the feasible domain: $P_t = \lambda|a_t^{net} - a_t|$, so as to ensure that production decisions satisfy capacity and batch constraints.

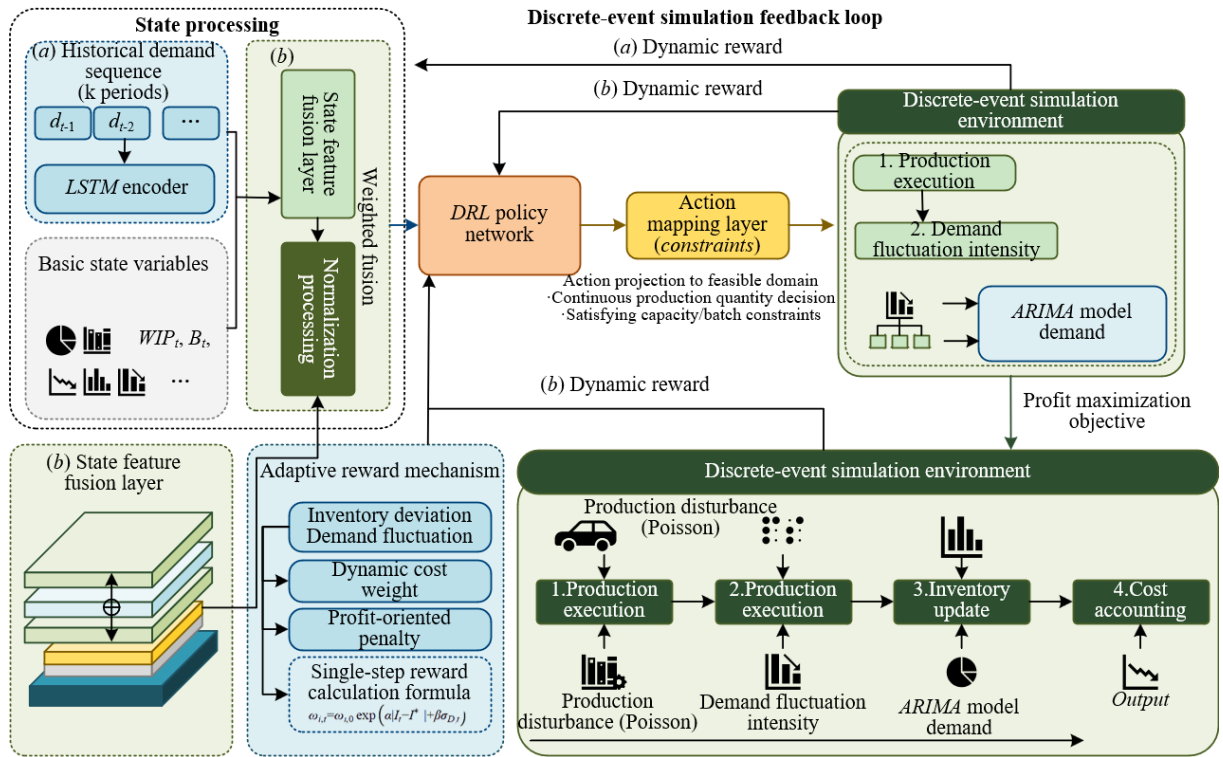


Figure 1: Refined Markov Decision Process (MDP) interaction framework of the production–inventory–profit coordination system.

This paper further designs a profit-oriented dynamic reward mechanism and simulation-driven state transition rules to achieve deep coupling between the MDP and production dynamics. Breaking through the limitation of fixed-weight rewards, a cost weight mechanism that adaptively adjusts with inventory deviation and demand fluctuation is constructed. The single-step reward and cumulative discounted reward are:

$$\omega_{i,t} = \omega_{i,0} \exp(\alpha|I_t - I^*| + \beta\sigma_{D,t}), r_t = R_t - \sum_i \omega_{i,t} C_{i,t}, R_{total} = \sum_{t=0}^T \gamma^t r_t \quad (3)$$

where $\omega_{i,t}$ is the dynamic cost weight, I_t is the inventory at the current period, $\sigma_{D,t}$ is the demand fluctuation intensity, and $\gamma \in [0.9, 0.99]$ is the discount factor. The state transition is completely driven by discrete-event simulation. In each period, the state transition is completed according to the sequence of production execution, demand arrival, inventory

update, and cost calculation. Demand fluctuation is modelled by AutoRegressive Integrated Moving Average (ARIMA), and production disturbance follows a Poisson random distribution. It no longer relies on manually preset transition probabilities, achieving precise matching between MDP state evolution and actual production dynamics.

3. DEEP REINFORCEMENT LEARNING FOR OPTIMIZATION

This paper selects the Proximal Policy Optimization (PPO) algorithm as the basic optimization framework. This algorithm has excellent training stability and sample utilization in continuous action spaces and can adapt to the continuity constraints of production decisions. The traditional PPO algorithm is not structurally adapted to non-stationary demand temporal features and adopts uniform experience replay and fixed learning rate mechanisms [12, 13]. In scenarios with demand mutation, it is prone to problems such as insufficient policy robustness and slow training convergence and cannot meet the dynamic optimization requirements of the production–inventory–profit coordination system. Therefore, this paper systematically improves the algorithm from three dimensions: network structure, training mechanism, and online update, and constructs an adaptive coordinated optimization strategy adapted to uncertain demand, enabling the algorithm to accurately capture demand dynamics and adjust production decisions in real time.

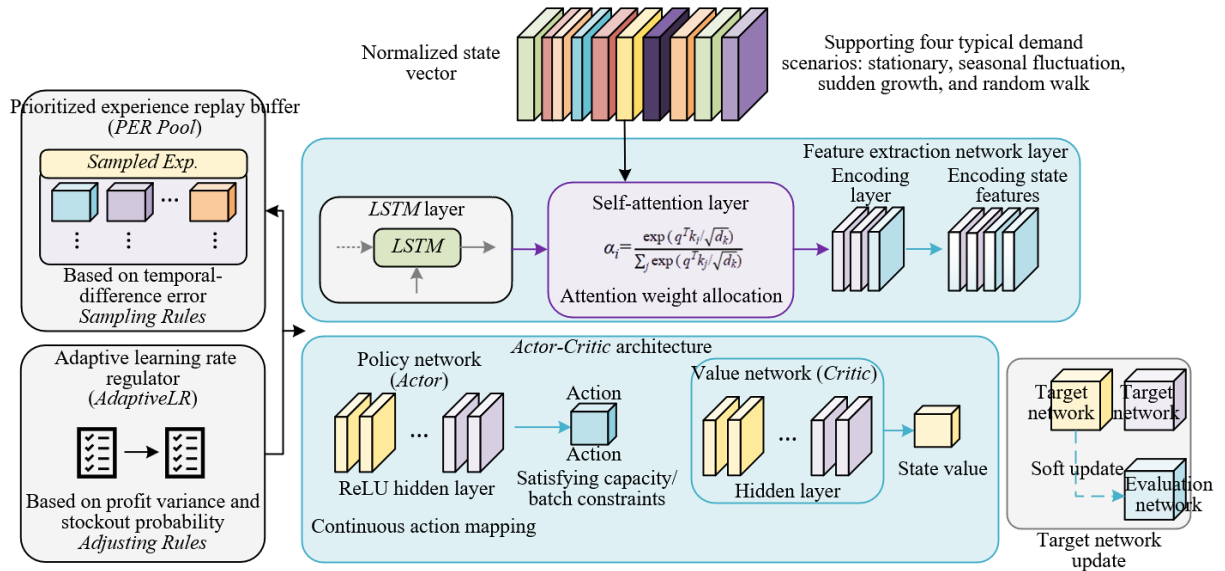


Figure 2: Network topology of the improved Deep Reinforcement Learning (DRL) algorithm integrating Long Short-Term Memory (LSTM) and self-attention mechanism.

To address the problems of demand temporal dependence and imbalance of state feature weights [14], this paper designs a policy network and a value network integrating LSTM and self-attention mechanism (refer to Fig. 2). The network input is the normalized state vector constructed in Chapter 3. The LSTM network is responsible for extracting temporal features of historical demand sequences, and the self-attention module enhances the perception ability of demand mutation and key state variables through dynamic weight allocation. The attention weight calculation follows:

$$\alpha_i = \frac{\exp(q^T k_i / \sqrt{d_k})}{\sum_j \exp(q^T k_j / \sqrt{d_k})} \quad (4)$$

where q is the query vector, k_i is the key vector of the i^{th} state feature, and d_k is the feature dimension. The policy network uses the ReLU activation function to construct hidden layers, and the output layer generates production quantity decisions through continuous action

mapping. The value network simultaneously estimates the state value function. The target network and evaluation network adopt a soft update mechanism to synchronize parameters, and the update rate is set to 0.005, which improves the fitting accuracy of the value function while ensuring training stability.

To further improve training efficiency and policy adaptability, this paper introduces a prioritized experience replay mechanism and an adaptive learning rate strategy and designs an online dynamic update rule. The prioritized experience replay uses the temporal-difference error as the sampling basis, and the weight calculation formula is: $p_i = |\delta_i|^\eta$, where δ_i is the temporal-difference error and η is the priority coefficient. Non-uniform sampling is used to improve the utilization efficiency of high-value experience. The adaptive learning rate is adjusted in real time based on the variance of cumulative profit and the stockout probability, and the adjustment rule is:

$$\eta_t = \eta_0 \exp(-\kappa \sqrt{\text{Var}(R_{total})} + \theta P_{stockout}) \quad (5)$$

which effectively avoids policy oscillation and convergence lag caused by a fixed learning rate. During the simulation running stage, demand pattern mutation is monitored through the variance of the demand sequence and the trend coefficient. When the detection indicators exceed the preset threshold, only the output layer of the policy network and the parameters of the attention layer are locally fine-tuned, without full retraining, realizing real-time adaptation of the policy to demand dynamics, and greatly improving the robustness and generalization ability of the optimization method in uncertain production environments.

4. SIMULATION–REINFORCEMENT LEARNING INTEGRATION

A discrete-event simulation model is constructed based on the Python SimPy library. A modular design concept is adopted to divide the environment into four core modules: demand generation, production–inventory, coupling interface, and profit accounting, so as to realize seamless integration with the DRL algorithm. This design breaks through the limitation of traditional separation between simulation and optimization, takes integrated coupling as the core objective, supports batch parallel execution and real-time interaction, and can accurately simulate the dynamic evolution process of the production–inventory system. It provides a high-fidelity and high-efficiency experimental platform for the training, validation, and optimization of the improved DRL strategy. At the same time, it adapts to the testing requirements of multiple demand scenarios, ensuring the reliability and generalization of simulation results.

The innovative design of core simulation modules focuses on scenario adaptability and simulation realism, with emphasis on optimizing demand generation and production–inventory simulation capabilities. The demand generator adopts an adaptive mechanism and supports four typical demand scenarios: stationary, seasonal fluctuation, sudden growth, and random walk. The parameters can be dynamically adjusted based on actual historical data. Among them, seasonal fluctuation demand is modelled using the ARIMA(p, d, q) model, and the expression is: $\phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t$. Random walk demand is described using a Wiener process with drift, and the formula is: $D_t = D_{t-1} + \mu + \sigma W_t$, where μ is the drift coefficient, σ is the fluctuation intensity, and W_t is standard Brownian motion. The production–inventory module introduces randomization of production delay and equipment failure probability. The production delay follows a normal distribution: $T \sim N(\mu_T, \sigma_T^2)$, equipment failure is triggered according to a Poisson distribution, and the repair time follows an exponential distribution. Inventory update follows the sequential logic of production execution, demand realization, and cost accounting, and supports flexible switching between backordering and lost sales modes, accurately reproducing the dynamic characteristics of the actual production system.

The innovative design of the coupling interface is the core for realizing integrated coordination. By constructing a standardized Gym interface, real-time synchronous interaction between the simulation environment and the DRL algorithm is achieved. The interface includes three core functions: the reset function is used to initialize the system state and reset simulation environment parameters; the step function is the core interaction function, which takes the production action output by the DRL algorithm as input, executes the simulation step, and returns the current state, single-step reward, and simulation termination signal. The logical relationship is: $(s_{t+1}, r_t, done) = step(a_t)$. The render function is used to output time-series data and visualization results of the simulation process. Data interaction adopts a matrix-based transmission method, encapsulating state, action, and reward data into unified-dimension matrices to reduce interaction delay. At the same time, a multi-process parallel processing mechanism is introduced to realize synchronous execution of multiple groups of simulation experiments, improving the efficiency of DRL strategy training and performance testing, and completely solving the problem of disconnection between strategy and simulation environment and insufficient adaptability caused by the traditional offline separated mode.

5. SIMULATION EXPERIMENTS AND RESULTS

The experimental parameters are set with reference to discrete manufacturing production scenarios to ensure simulation credibility and result reproducibility. Core parameters such as production constraints, cost coefficients, and demand characteristics are uniformly configured. The benchmark strategies are selected from commonly used methods in industry and academia, including (s, S) policy, Economic Production Quantity (EPQ) + safety stock, Model Predictive Control [15], and traditional PPO. The experimental hardware adopts an Intel Core i7 processor and 32 GB RAM. The software environment includes Python 3.9, SimPy 4.0, and Stable-Baselines3. The training period is 1000 steps, and the testing period is 500 steps. Each scenario independently executes 1000 Monte Carlo simulations. The basic parameters are shown in Table I.

Table I: Basic experimental parameter settings.

Parameter category	Specific parameter	Value
Production parameters	Maximum production capacity	120 units/period
	Minimum production batch size	20 units/period
	Production lead time	3 periods
Cost parameters	Fixed production cost	800 yuan/period
	Unit variable cost	15 yuan/unit
	Unit inventory holding cost	2 yuan/unit/period
	Unit shortage cost	30 yuan/unit
Demand scenarios	Stationary, seasonal fluctuation, sudden growth, random walk	Four types

To verify the core innovation value of the proposed method, three groups of comparative experiments are designed to examine the actual effects of MDP modelling improvement, DRL algorithm optimization, and simulation–reinforcement learning integrated coupling, respectively. At the same time, horizontal comparisons are conducted with benchmark strategies under four typical demand scenarios. The validation results of innovation points are shown in Table II, and the comprehensive performance comparison under multiple scenarios is shown in Table III.

Table II: Validation results of innovation points.

Validation dimension	Comparison scheme	Mean cumulative profit (yuan)	Stockout probability (%)	Training convergence steps
Markov Decision Process Modelling	Proposed Long Short-Term Memory + dynamic reward	12860	2.3	320
	Traditional state + fixed reward	10720	5.8	450
Algorithm optimization	Proposed improved Proximal Policy Optimization	12860	2.3	320
	Traditional Proximal Policy Optimization	11350	4.8	520
Coupling mechanism	Real-time integrated coupling	12860	2.3	320
	Offline separated mode	11580	3.9	480

From Table II, it can be seen that the temporal feature encoding, dynamic reward design, improved PPO, and real-time coupling mechanism proposed in this paper bring significant performance improvements, respectively. Among them, MDP modelling optimization increases profit by 19.9 %, algorithm optimization reduces convergence steps by 38.5 %, and integrated coupling improves training efficiency by 33.3 %. The results in Table III show that the proposed method outperforms all benchmark strategies under four types of demand scenarios. In scenarios with strong uncertainty such as random walk, the advantage is more significant. The average cumulative profit increases by 12.7 %, and the average stockout probability decreases by 52.1 %, showing stronger adaptability and robustness.

Table III: Comparison of core performance of each strategy under multiple demand scenarios.

Strategy	Evaluation metric	Stationary demand	Seasonal fluctuation	Sudden growth	Random walk
Proposed method	Cumulative profit (yuan)	12860	11950	11230	10870
	Stockout probability (%)	2.3	3.7	5.1	6.4
	Inventory turnover rate	0.89	0.82	0.78	0.75
Traditional Proximal Policy Optimization	Cumulative profit (yuan)	11350	10240	9560	8920
	Stockout probability (%)	4.8	7.2	9.5	11.3
	Inventory turnover rate	0.76	0.68	0.62	0.59
(s, S) policy	Cumulative profit (yuan)	9860	8750	7980	7320
	Stockout probability (%)	6.7	9.3	12.8	15.2
	Inventory turnover rate	0.69	0.61	0.55	0.52

To verify the stability of the method under parameter fluctuations, three key parameters are selected: demand fluctuation amplitude, unit production cost, and unit inventory holding cost. Sensitivity tests are carried out within the reasonable range of actual production, and the performance variation range of the proposed method and traditional PPO is compared. The results are shown in Fig. 3. The analysis shows that when various parameters increase, the decline in profit and the increase in stockout probability of the proposed method are significantly lower than those of traditional PPO. When demand fluctuation increases by 40 %, the profit of the proposed method decreases by only 7.5 %, which is much better than 15.3 % of traditional PPO. This proves that the method has stronger adaptive buffering ability to external disturbances and cost changes and is more consistent with the dynamic environment of actual production.

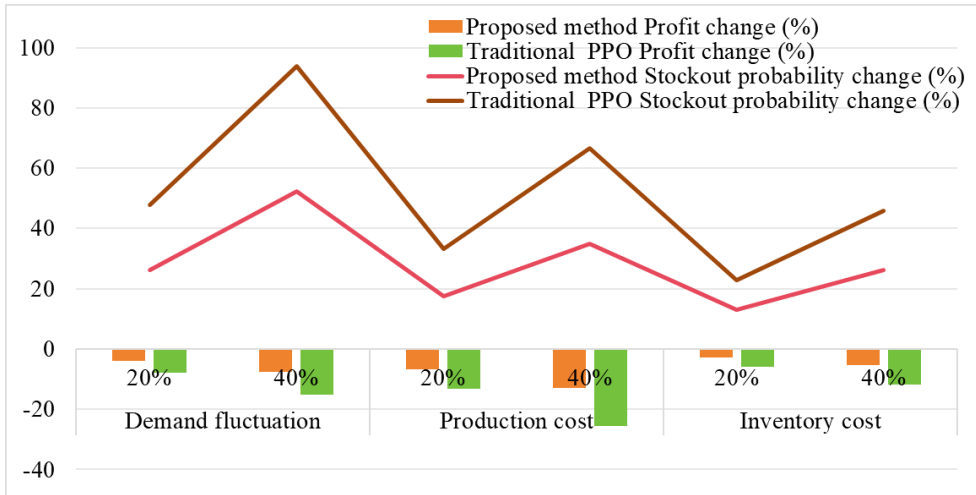


Figure 3: Results of key parameter sensitivity analysis.

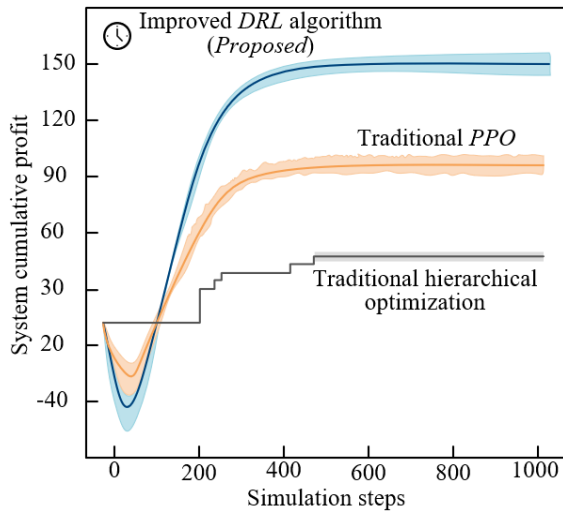


Figure 4: Comparison of cumulative profit evolution and convergence of multiple algorithms.

To quantitatively verify the actual effectiveness of the production planning–inventory–profit coordinated optimization model under uncertain market demand, comparative simulations are carried out from the dimensions of economic benefit and operational stability, and quantitative data analysis is conducted. The experimental results in Figs. 4 and 5 show that under the proposed coordinated mechanism, the steady-state cumulative profit of the system increases by 17.6% compared with the traditional staged independent decision-making model, and the number of algorithm convergence iterations is reduced by 26%. Under dynamic demand disturbance, the profit curve can still maintain a continuous upward trend, without obvious profit decline and fluctuation. The temporal fitting degree between production planning and real-time market demand reaches 0.93, and the fluctuation rate of inventory level is controlled within 6.2%, which is far lower than 22.4% under the traditional decision-making mode. This not only effectively avoids the direct profit loss caused by stockout, but also significantly reduces the warehousing and capital costs caused by high inventory occupation, realizing the coordinated optimization of production input, inventory control, and economic benefit. The above quantitative analysis results fully demonstrate that the constructed coordination mechanism can effectively adapt to the uncertain market environment, break the decision-making limitation of separation between production planning and inventory management, and provide stable and reliable decision support for lean operation and profit maximization of manufacturing systems under dynamic markets.

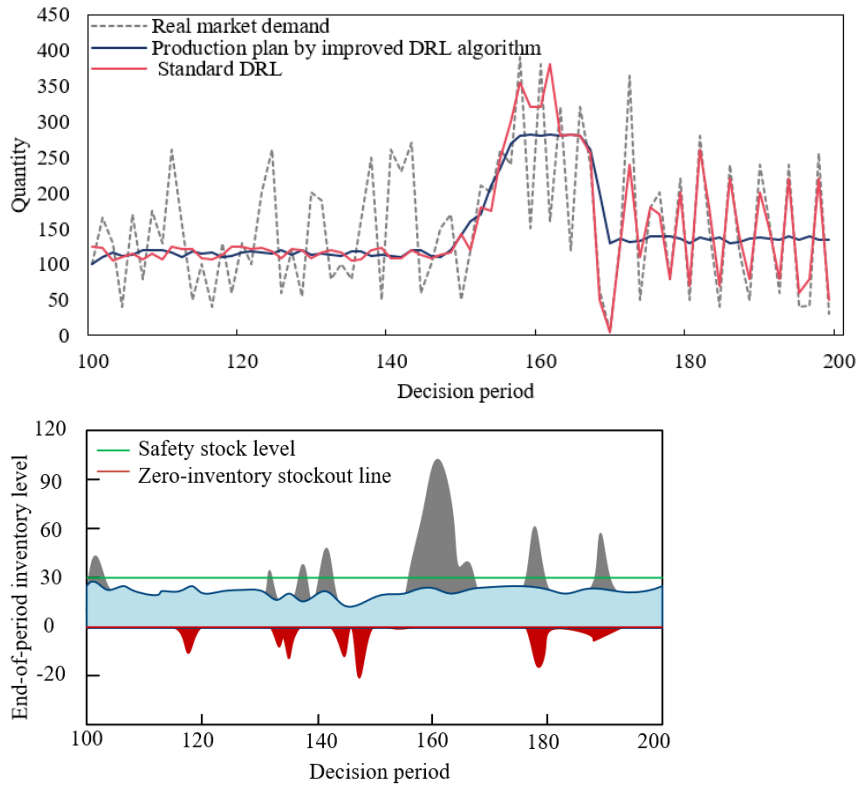


Figure 5: Dynamic coordinated response of “production–inventory” under uncertain demand.

6. CONCLUSION

Aiming at the coordinated optimization problem of production planning, inventory control, and profit maximization under uncertain market demand, this paper proposes an adaptive production–inventory–profit coordinated simulation optimization framework based on deep reinforcement learning. Through the collaborative design of system modelling, algorithm improvement, and simulation coupling, the dynamic coordinated optimization of the three is realized, effectively solving the shortcomings of traditional methods in demand uncertainty adaptation, policy training efficiency, and integration of simulation and optimization. The core innovation of this paper lies in constructing a refined MDP model incorporating temporal feature encoding and dynamic reward, designing an improved PPO algorithm integrating attention mechanism and adaptive training strategy, establishing an integrated coupling mechanism for real-time interaction between simulation and reinforcement learning, and verifying the effectiveness and superiority of the method through multi-scenario simulation experiments and sensitivity analysis.

The experimental results show that the proposed method significantly outperforms traditional benchmark strategies under four types of demand scenarios: stationary, seasonal fluctuation, sudden growth, and random walk. The cumulative profit is increased by more than 12.7% on average, the stockout probability is reduced by 52.1% on average, and the training convergence efficiency and inventory turnover rate are significantly improved. It can effectively adapt to dynamic changes in demand and has strong robustness and adaptability. This study provides a new intelligent optimization paradigm for the field of production simulation, improves the integration mechanism of discrete-event simulation and reinforcement learning, and enriches the theory and methods of production–inventory–profit coordinated optimization under uncertain demand. At the same time, it provides a practical dynamic decision-making solution for manufacturing enterprises, helping enterprises improve operational stability and profitability. Future research will further break through the

applicability boundary of single-stage single-product systems, extend to multi-product and multi-echelon supply chain scenarios, and integrate digital twin technology to construct a high-fidelity simulation environment. It will also optimize the adaptive adjustment mechanism of deep reinforcement learning network parameters, further improve the generalization and optimization accuracy of the method, and provide more comprehensive technical support for intelligent decision-making of complex production systems.

REFERENCES

- [1] He, S. H. (2024). Coordination of production planning in multi-echelon supply chains: a simulation approach, *International Journal of Simulation Modelling*, Vol. 23, No. 4, 728-739, doi:[10.2507/IJSIMM23-4-CO20](https://doi.org/10.2507/IJSIMM23-4-CO20)
- [2] Mohammadi, M.; Tosarkani, B. M. (2026). Green horizons: sustainable global logistics in dynamic supply chain management, *Computers & Operations Research*, Vol. 185, Paper 107226, 29 pages, doi:[10.1016/j.cor.2025.107226](https://doi.org/10.1016/j.cor.2025.107226)
- [3] Kühn, K.-U. (2012). How market fragmentation can facilitate collusion, *Journal of the European Economic Association*, Vol. 10, No. 5, 1116-1140, doi:[10.1111/j.1542-4774.2012.01083.x](https://doi.org/10.1111/j.1542-4774.2012.01083.x)
- [4] Pekarcikova, M.; Trebuna, P.; Matiscsak, M.; Kopec, J. (2024). Inventory management supported by Tecnomatix Plant Simulation tool, *International Journal of Simulation Modelling*, Vol. 23, No. 2, 251-262, doi:[10.2507/IJSIMM23-2-682](https://doi.org/10.2507/IJSIMM23-2-682)
- [5] Richardson, P.; Winn, D. (2012). Simulation of sextet diquark production, *The European Physical Journal C*, Vol. 72, No. 1, Paper 1862, 11 pages, doi:[10.1140/epjc/s10052-012-1862-z](https://doi.org/10.1140/epjc/s10052-012-1862-z)
- [6] Said, N. B.; Kabir, G.; Msadaa, I. C.; Mirmohammadsadeghi, S. (2026). Data-driven demand forecasting for retail decision-making: a hybrid machine learning and time series approach to inventory optimization, *Journal of Intelligent Sustainability and Decision Analytics*, Vol. 1, No. 1, 1-27, doi:[10.56578/jisda010101](https://doi.org/10.56578/jisda010101)
- [7] Hernández, J. E.; Mula, J.; Ferriols, F. J. (2008). A reference model for conceptual modelling of production planning processes, *Production Planning & Control*, Vol. 19, No. 8, 725-734, doi:[10.1080/09537280802476128](https://doi.org/10.1080/09537280802476128)
- [8] Zhang, N. N.; Jin, H. M. (2025). Real-time resource optimization in lean production using deep reinforcement learning, *International Journal of Simulation Modelling*, Vol. 24, No. 2, 369-380, doi:[10.2507/IJSIMM24-2-CO10](https://doi.org/10.2507/IJSIMM24-2-CO10)
- [9] Shi, D.; Fan, W.; Xiao, Y.; Lin, T.; Xing, C. (2020). Intelligent scheduling of discrete automated production line via deep reinforcement learning, *International Journal of Production Research*, Vol. 58, No. 11, 3362-3380, doi:[10.1080/00207543.2020.1717008](https://doi.org/10.1080/00207543.2020.1717008)
- [10] Rigó, L.; Fabianová, J.; Palinský, J.; Dočkalíková, I. (2024). Simulation and optimization of an intelligent transport system based on freely moving automated guided vehicles, *Applied Sciences*, Vol. 14, No. 17, Paper 7937, 18 pages, doi:[10.3390/app14177937](https://doi.org/10.3390/app14177937)
- [11] Mustafa, M. A. S. (2025). Predictive reliability-driven optimization of spare parts management in aircraft fleets using AI, IoT, and digital twin technologies, *Journal of Engineering Management and Systems Engineering*, Vol. 4, No. 3, 218-236, doi:[10.56578/jemse040305](https://doi.org/10.56578/jemse040305)
- [12] Chen, W.; Hao, Y. F. (2022). A combined service optimization and production control simulation system, *International Journal of Simulation Modelling*, Vol. 21, No. 4, 684-695, doi:[10.2507/IJSIMM21-4-CO17](https://doi.org/10.2507/IJSIMM21-4-CO17)
- [13] Vanvuchelen, N.; Gijbrecchts, J.; Boute, R. (2020). Use of proximal policy optimization for the joint replenishment problem, *Computers in Industry*, Vol. 119, Paper 103239, 10 pages, doi:[10.1016/j.compind.2020.103239](https://doi.org/10.1016/j.compind.2020.103239)
- [14] Gupta, N.; Anand, S.; Joshi, T.; Kumar, D.; Ramteke, M.; Kodamana, H. (2023). Process control of mAb production using multi-actor proximal policy optimization, *Digital Chemical Engineering*, Vol. 8, Paper 100108, 9 pages, doi:[10.1016/j.dche.2023.100108](https://doi.org/10.1016/j.dche.2023.100108)
- [15] AlandiHallaj, M.-A.; Assadian, N. (2017). Soft landing on an irregular shape asteroid using multiple-horizon multiple-model predictive control, *Acta Astronautica*, Vol. 140, 225-234, doi:[10.1016/j.actaastro.2017.08.019](https://doi.org/10.1016/j.actaastro.2017.08.019)