Optimization of Inventory Allocation Strategies for Cross-Border E-Commerce Based on D3S-CI Model: A Combination of Multi-Objective and Multi-Stage Approaches

Sangbing Tsai[1*], Jaheer Mukthar.KP [2]

[1]International Engineering and Technology Institute, Hong Kong; klj0418@gmail.com

[2]Kristu Jayanti College Autonomous Bengaluru, India; jaheermukthar@gmail.com

*Corresponding Author: klj0418@gmail.com

DOI: https://doi.org/10.30209/IJMO.202503.017

Submitted: Oct. 10, 2025 Accepted: Nov. 26, 2025

ABSTRACT

Effective inventory control remains a formidable challenge in cross-border e-commerce supply chains, owing to significant demand volatility and the necessity for multifunctional optimization. In response, this paper presents D3S-CI, an intelligent inventory management framework that synthesizes deep learning techniques with adaptive decision-making processes. The proposed framework employs multi-source data to dynamically recalibrate inventory policies, thereby optimizing distribution networks and concurrently curtailing operational costs and stockout incidence. Empirical evaluations indicate that D3S-CI yields superior outcomes compared to prevailing state-of-the-art benchmarks, attaining an aggregate performance enhancement of 9%-12%. Ablation analyses corroborate the indispensability of the model's constituent state representation and decision-making modules. The demonstrated resilience of the model in fluctuating market conditions indicates considerable potential for its practical implementation.

Keywords: Deep reinforcement learnin, Cross-border e-commerce, Inventory management, Inventory optimization, Dynamic adjustment, Data fusion

1. Introduction

With the rapid expansion of global e-commerce, cross-border supply chains are encountering increasingly complex operational challenges. As a pivotal element of supply chain management, inventory control directly influences operational efficiency, cost management, and customer satisfaction [1]. Inventory management in cross-border e-commerce necessitates real-time support and must balance multiple objectives-such as inventory costs, stock levels, and delivery timelines-within a highly dynamic, complex, and uncertain environment. Consequently, formulating effective inventory allocation strategies has become a critical determinant for achieving success in the competitive cross-border e-commerce landscape.

However, existing inventory management approaches face several limitations. Traditional models, such as the Economic Order Quantity (EOQ), presume stable and known demand, rendering them inadequate for handling the significant demand fluctuations typical in cross-border e-commerce [2]. The ABC classification method improves inventory management in certain aspects by optimizing resource allocation through inventory categorization; however, it fails to address the multi-objective optimization of inventory costs, stock levels, and lead times in an integrated manner [3]. Time series-based demand forecasting methods are widely employed to project future demand from historical data. Nevertheless, their reliance on static datasets often diminishes flexibility in responding to dynamic supply chain changes [4]. Furthermore, heuristic algorithms like Genetic Algorithms (GA) and Simulated Annealing (SA) are applied to inventory optimization problems. While these methods seek optimal solutions to reduce holding costs or improve stock utilization, they struggle with the complex, multi-objective challenges of inventory distribution [5]. Additionally, such methods are typically incapable of optimizing inventory allocation across multiple time scales, thereby undermining supply chain adaptability under volatile conditions.

To address these issues, this paper proposes D3S-CI (Deep Reinforcement Learning-based Dynamic Supply Chain Inventory Allocation), a model designed to optimize inventory allocation strategies for cross-border e-commerce. The D3S-CI model integrates advanced deep learning, multi-objective optimization, and a multi-step decision mechanism, enabling dynamic inventory allocation in complex cross-border e-commerce environments while simultaneously optimizing multiple goals, including inventory costs, stock levels, and shipping efficiency. By incorporating short-term, medium-term, and long-term optimization strategies, the D3S-CI model can tailor inventory policies to different temporal scales, effectively responding to market volatility and supply chain disruptions [6]. The main contributions of this paper are as follows:

• We propose D3S-CI, a model based on deep reinforcement learning that integrates multi-objective optimization and a multi-step decision mechanism to ensure dynamic and adaptive inventory allocation in cross-border e-commerce supply chains.

• We design a multi-stage response mechanism that optimizes inventory distribution across short-term, medium-term, and long-term horizons, thereby enhancing the system's flexibility and adaptability.

• We establish hybrid incremental learning and causal sequence intervention mechanisms to strengthen the model's ability to accommodate sudden market shifts and improve its long-term stability.

The structure of this paper is organized as follows: Section 2 reviews the advancements and shortcomings of existing inventory management methods, with a particular focus on multi-objective and multi-stage optimization problems. Section 3 provides a detailed description of the D3S-CI model, including the functionality and architecture of its state representation, basic decision-making, and adaptive modules. Section 4 presents the experimental validation of the model, covering the datasets, experimental setup, evaluation metrics, and comparative experiments, which demonstrate the advantages of D3S-CI in cross-border e-commerce inventory optimization. Finally, Section 5 concludes the paper and discusses potential future research directions and improvements.

2. Related Work

2.1 Traditional Inventory Optimization Methods

Traditional inventory optimization methods have achieved considerable success in various supply chain management applications. Notable examples include approaches based on Linear Programming (LP) and Integer Programming (IP). LP optimizes inventory allocation by formulating the inventory management model as a set of linear mathematical expressions and is widely used in multi-warehouse and multi-product inventory optimization [7]. Another established method is the profit-maximization model, which determines the optimal order quantity and replenishment timing by considering holding costs, stockout costs, and ordering costs; it is often effective under stable demand conditions [8]. Furthermore, Dynamic Programming (DP) is frequently employed to address multi-stage decision-making in inventory management, particularly in long-term strategy planning, aiding decision-makers in identifying optimal policies over time [9]. However, these methods often entail substantial computational resources and rely on specific assumptions, rendering them less adaptable to the complex and volatile market environment of cross-border e-commerce. Fuzzy Logic Control (FLC), another classical technique, models uncertainty and ambiguity using fuzzy sets and reasoning rules, making it suitable for inventory management under uncertain demand. Nevertheless, its effectiveness can be limited in highly dynamic and complex markets [10]. Finally, Gray System Theory is utilized to build forecasting models that handle both known and unknown information, applicable in situations where demand and supply chain states are difficult to predict accurately. Yet, its application in multi-objective optimization remains relatively limited [11].

In contrast, the proposed D3S-CI model offers distinct advantages over these traditional methods. Unlike conventional models, D3S-CI integrates deep learning to dynamically adjust inventory allocation based on real-time demand shifts and supply chain fluctuations. Moreover, it simultaneously optimizes multiple inventory management objectives, such as cost, warehouse utilization rates, and transportation efficiency.

2.2 Inventory Optimization Method Based on Deep Learning

Convolutional Neural Networks (CNNs) are employed to extract features from historical data and to predict inventory demand and supply patterns. They are particularly valued for their advantages in processing image-like and multi-dimensional data [12]. Another prominent approach involves the use of Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, which are distinguished by their capability to process time-series data and effectively forecast future demands, making them highly suitable for large-scale scenarios [13]. Additionally, Autoencoders (AEs) are utilized for dimensionality reduction and feature extraction from inventory data, thereby aiding in the optimization of inventory allocation decisions, particularly when dealing with large datasets [14]. Generative Adversarial Networks (GANs) are also applied to generate synthetic data, enhance training datasets, and ultimately improve the robustness of inventory predictions [15]. Finally, Reinforcement Learning (RL) has been incorporated into deep learning frameworks for inventory management, where it progressively learns optimal policies through interaction with the inventory management environment [16]. However, these methods often struggle to simultaneously enhance prediction accuracy, optimize strategies for multiple objectives, and respond swiftly to market evolution within the complex context of cross-border e-commerce.

In contrast to other deep learning-based inventory optimization methods, the proposed D3S-CI model integrates multi-step optimization, enabling it to address short-term, medium-term, and long-term objectives concurrently. This integration significantly enhances its adaptability and flexibility under the dynamic and complex conditions of e-commerce supply chains.

2.3 The Application of RL in Inventory Management

There is growing attention focused on the application of reinforcement learning (RL) in inventory management. Classical RL algorithms, such as Q-Learning, are widely used to optimize inventory levels and identify optimal policies for replenishment decisions [17]. However, Q-Learning faces the challenge of the exponential growth of state spaces, making it difficult to effectively manage large-scale inventory problems. To address this issue, Deep Q-Networks (DQN) integrate deep neural networks to mitigate the problems associated with high-dimensional state spaces [18]. Beyond value-based methods, policy gradient algorithms, such as Proximal Policy Optimization (PPO) and Asynchronous Advantage Actor-Critic (A3C), are also employed. These methods are particularly suitable for dynamic and complex inventory environments [19]. Furthermore, multi-agent reinforcement learning (MARL) is applied to inventory management systems involving multiple warehouses or suppliers, where multiple agents collaborate to optimize overall inventory decisions and enhance management efficiency [20]. Notably, PPO is widely adopted in inventory management due to its ability to ensure stable and consistent training performance while adapting to uncertain demand and dynamic conditions [21].

Unlike these traditional RL methods, the proposed D3S-CI model integrates reinforcement learning with multi-objective optimization to balance various goals rather than focusing on a single objective. It establishes a multi-stage optimization mechanism to address both short-term fluctuations and long-term changes in the cross-border e-commerce environment, thereby significantly enhancing flexibility and adaptability. Furthermore, D3S-CI dynamically adjusts inventory strategies through continuous interaction with the environment, enabling steady policy improvement. This adaptive learning capability allows the model to provide more comprehensive and robust solutions for inventory optimization.

3. Method

3.1. Overview of Our Model

The D3S-CI model is designed to optimize dynamic inventory allocation strategies within cross-border e-commerce supply chains by integrating deep learning enhancement, multi-objective optimization, and multi-stage optimization mechanisms. Figure 1 illustrates the overall architecture of the model and its constituent modules. Each module operates efficiently, from input data processing to final inventory distribution decisions, enabling the model to respond in real-time to market demands and supply chain fluctuations.

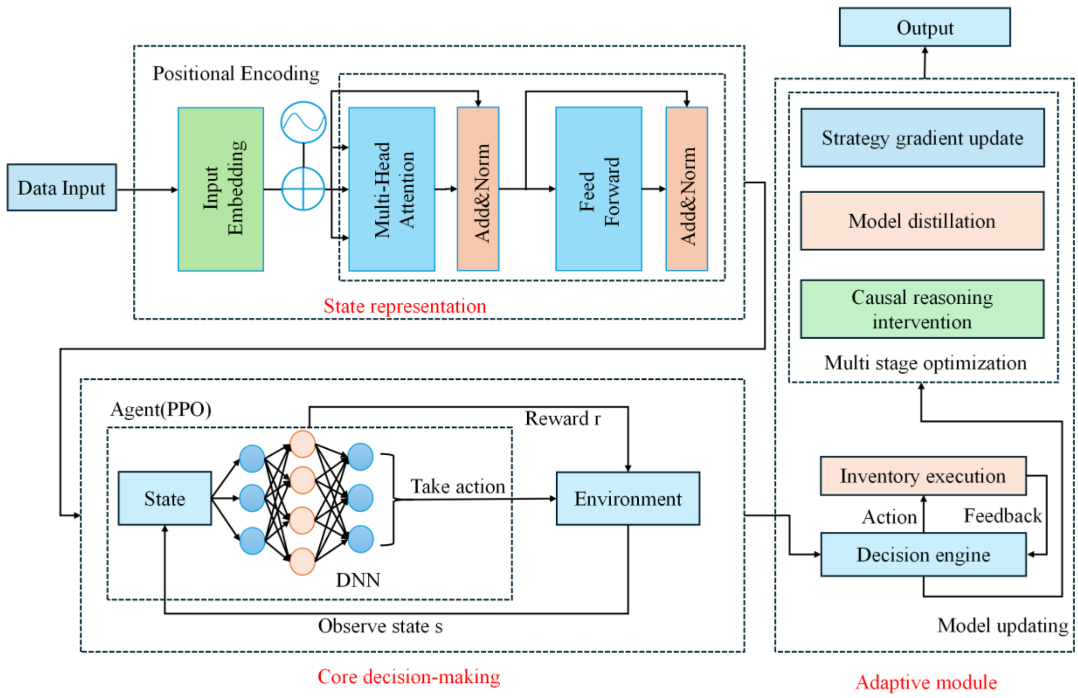

Figure 1. Overall architecture of the d3s-ci model for dynamic inventory allocation and multi-objective optimization in cross-border e-commerce supply chain.

The model ingests multi-dimensional data from the cross-border e-commerce supply chain, including inventory levels, order demands, logistics information, and supplier lead times [22]. These data are preprocessed by the state representation module to extract salient features from the complex raw inputs. This module not only cleans and normalizes the raw data but also integrates them through feature engineering techniques, ensuring input quality and consistency. The preprocessed data are subsequently fed into a Transformer model for advanced feature extraction. Utilizing a self-attention mechanism, the Transformer effectively captures temporal dependencies and global patterns within the data, generating a comprehensive state vector with an enhanced representation.

At the core of the model, the Deep Reinforcement Learning (DRL) module plays a pivotal role. Employing RL algorithms such as PPO or SAC, this module optimizes inventory allocation strategies through continuous interaction with the environment [23]. The DRL agent dynamically adjusts the inventory distribution plan in response to market changes, thereby maximizing the overall efficiency of the supply chain. Particularly under demand volatility or supply chain disruptions in cross-border e-commerce, the DRL module can facilitate timely and optimal decision-making. Furthermore, the model incorporates multi-objective optimization to ensure a balanced consideration of various goals, such as inventory costs, stockout rates, and transportation efficiency.

The adaptive module enables continuous optimization of inventory distribution across different time scales through hybrid incremental learning and a multi-step optimization mechanism. In the short term, policy gradient methods update the model's parameters to respond swiftly to immediate demand shifts. In the medium term, a model distillation approach is applied to generate lightweight student models, enhancing computational efficiency. For the long term, causal intervention techniques allow the model to adapt to evolving market trends and assess the prolonged impact of strategies. Through this adaptive learning framework, the inventory allocation strategy is perpetually refined based on real-time market feedback, endowing the model with high flexibility and adaptability.

In summary, the architectural design ensures that D3S-CI exhibits strong robustness within the complex and dynamic cross-border e-commerce supply chain environment. By synergistically integrating deep reinforcement learning, multi-objective optimization, and multi-stage optimization, the model dynamically optimizes inventory allocation in real-time under fluctuating market conditions, thereby improving overall supply chain performance and responsiveness. Concurrently, the system effectively handles multi-dimensional data and ensures seamless interoperability between modules via efficient task scheduling and coordination mechanisms, significantly enhancing the efficacy of supply chain management.

3.2. State Representation Module: Transformer-based Cross-Border E-Commerce Supply Chain State Modeling and Representation

In the D3S-CI model, the state representation module processes multi-dimensional cross-border e-commerce supply chain data using a Transformer architecture to generate state vectors that accurately reflect supply chain dynamics. The core objective of this module is to transform complex raw input data into machine-readable state representations, thereby providing high-quality input for subsequent inventory allocation decisions. Figure 2 illustrates the structure and workflow of this module. Leveraging the self-attention mechanism, the Transformer effectively captures long-term dependencies and complex feature interactions within the supply chain data, delivering precise state vectors to the core decision-making module.

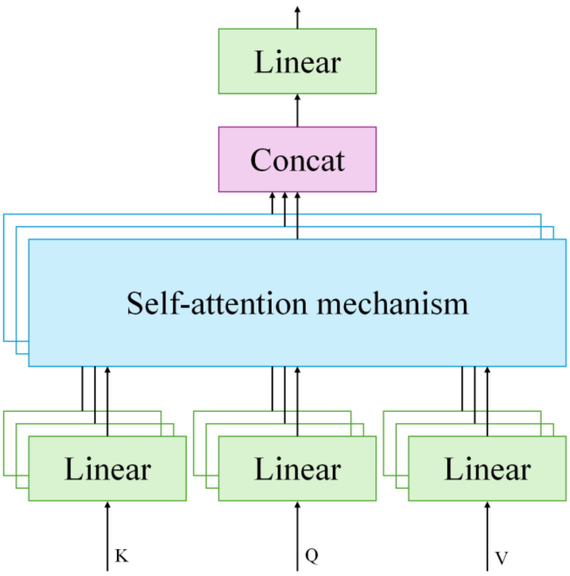

Figure 2. Architecture of the state representation module for inventory allocation in cross-border e-commerce supply chain.

Within the state representation module, input data first undergoes preprocessing and feature extraction to obtain feature x_t (e.g., inventory status, demand, and transportation information). Let μ and σ denote the mean and standard deviation of the input data, respectively. The data are standardized to $\hat{x}_t$ and scaled to a uniform range to facilitate subsequent processing, as shown in Equation (1).

$$\hat{x}_t = \frac{x_t - \mu}{\sigma} \cdots\cdots\cdots\cdots \text{[Formular 1]}$$

Following preprocessing, the data are fed into the Transformer model for state encoding. The Transformer captures dependencies across different time steps by computing the Query (Q), Key (K), and Value (V) matrices at each time step. Given the input data matrix X=[x_1,x_2,…,x_T], and letting W^Q, W^K, and W^V represent the weight matrices for the query, key, and value, respectively, the computation is defined in Equation (2).

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \cdots\cdots \text{[Formular 2]}$$

The self-attention mechanism dynamically adjusts attention weights based on the similarity between queries and keys, ultimately generating a comprehensive state representation s_t. d_k denotes the dimensionality of the key vector. The similarity QK^T determines the attention weights assigned to different parts of the input, thereby influencing the computation of the state representation, as formulated in Equation (3). This mechanism ensures that the most relevant features receive higher weights, enhancing the informativeness of the state vector.

$$s_t = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \cdots\cdots \text{[Formular 3]}$$

Subsequently, the generated state vector s_t is passed through a fully connected (FC) layer for further refinement, producing the final optimized state representation $\tilde{s}_t$. This step enables the model to extract higher-level features from the initial state representation, providing enriched input for the subsequent decision-making module, as expressed in Equation (4).

$$\tilde{s}_t = \text{FC}(s_t) \cdots\cdots\cdots\cdots \text{[Formular 4]}$$

Through this process, the Transformer effectively captures complex temporal patterns within the cross-border e-commerce supply chain, generating precise and informative state representations. This equips the D3S-CI model with robust data representation capabilities, ensuring accurate reflection of multi-dimensional information such as market demand, inventory status, and supply chain fluctuations. The state representation module thereby provides the necessary foundation for making more accurate and flexible inventory distribution decisions, enhancing the model's adaptability to dynamic changes.

3.3. Core Decision-Making Module: DRL-Driven Inventory Allocation Optimization in Cross-Border E-Commerce Supply Chains

The core decision-making module of the D3S-CI model employs a Deep Reinforcement Learning (DRL) algorithm to optimize inventory distribution strategies within cross-border e-commerce supply chains. Figure 3 illustrates the architecture of this module. Through continuous interaction with the environment, the module learns to dynamically adjust inventory allocation in response to market demand and supply chain fluctuations, thereby maximizing overall supply chain efficiency. This module not only addresses complex multi-objective optimization challenges but also enables real-time decision-making, ensuring both flexibility and efficiency in inventory management.

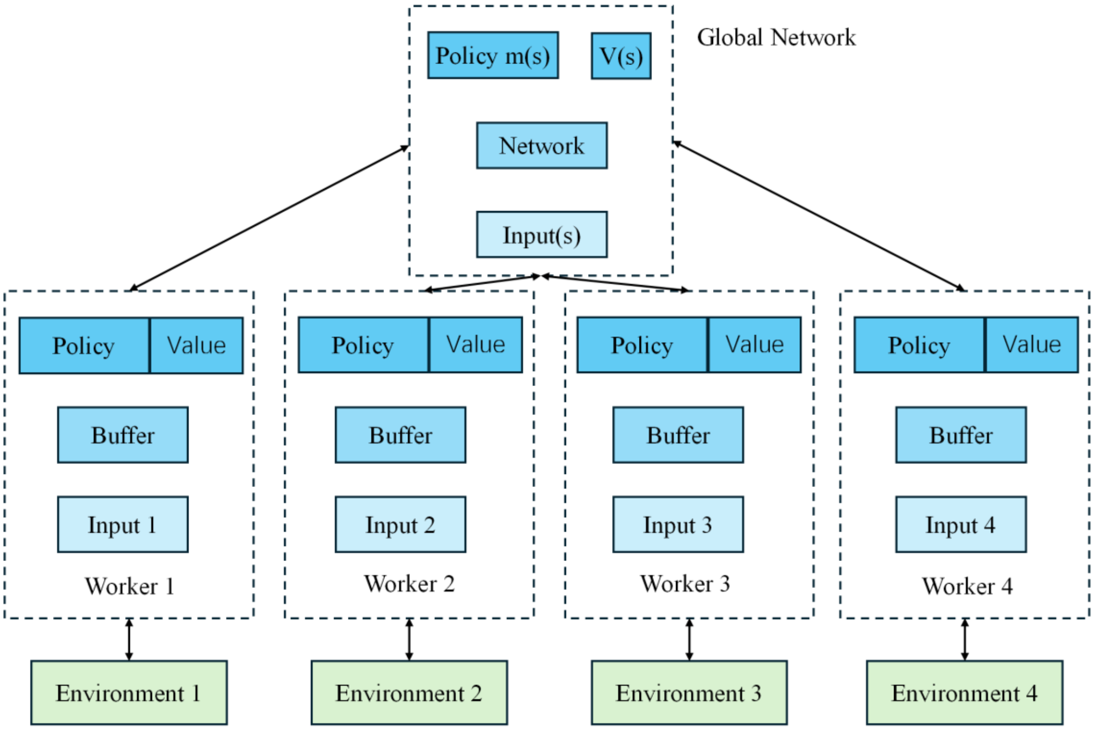

Figure 3. Architecture of the core decision module for inventory allocation optimization in cross-border e-commerce supply chain.

The fundamental objective of this module is to learn an optimal inventory distribution policy using the DRL framework. The state vector s_t, derived from the state representation module, captures the current supply chain status. The goal of reinforcement learning is to develop an optimal policy through environmental interaction to maximize long-term cumulative rewards (e.g., reducing holding costs and minimizing stockout rates). Through iterative updates in RL, the policy π is continuously refined, leading to improved inventory allocation. At each time step, the system selects an action a_t (e.g., replenishment, inventory redistribution) based on the current state s_t and the policy π, as defined in Equation (5).

$$a_t = \pi(s_t) \cdots\cdots \text{[Formular 5]}$$

After executing action atat, the system transitions to a new state st+1st+1 and receives an immediate reward r_t, subsequently updating the policy π. The immediate reward function $f(s_t, a_t)$ quantifies the inventory performance after taking a specific action, typically incorporating factors such as holding costs, stockout costs, and transportation expenses, as shown in Equation (6).

$$r_t = f(s_t, a_t) \cdots\cdots \text{[Formular 6]}$$

To optimize the policy, the Deep Q-Network (DQN) method is employed. DQN selects optimal actions by estimating the Q-value function $Q(s_t, a_t)$, where γ denotes the discount factor, representing the importance of future rewards. The Q-value function estimates the expected cumulative future reward when taking action a_t in state s_t. The model continually updates the Q-values to approximate the optimal policy, as formulated in Equation (7).

$$Q(s_t, a_t) = \mathbb{E}[r_t + \gamma \max_{a'} Q\ (s_{t+1}, a')] \cdots\cdots\cdots\cdots \text{[Formular 7]}$$

In the D3S-CI model, policy updates are performed using policy gradient methods to optimize the parameterized policy function. Through gradient ascent, the model learns policies that maximize long-term rewards. Here, θ_t represents the parameters of the policy network, α is the learning rate, and $\nabla_\theta \mathbb{E}[r_t]$ is the policy gradient, indicating the direction for adjusting policy parameters to maximize expected rewards, as expressed in Equation (8).

$$\theta_{t+1} = \theta_t + \alpha \nabla_\theta \mathbb{E}[r_t] \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 8]}$$

Through continuous policy iteration, the core decision-making module achieves optimal inventory allocation strategies for the supply chain. Unlike conventional methods, DRL simultaneously optimizes multiple objectives, such as reducing holding costs, minimizing stockouts, and improving transportation efficiency. More importantly, DRL adapts to dynamically changing environments and updates inventory allocation strategies in real-time. In summary, this deep reinforcement learning-based decision-making module dynamically adjusts inventory strategies, balances multiple objectives, and adapts to both market demands and the complex structure of cross-border e-commerce supply chains.

3.4. Adaptive Module: Hybrid Incremental Learning and Multi-Stage Optimization

The adaptive module of the D3S-CI model facilitates continuous optimization of inventory distribution strategies in response to dynamic changes and market uncertainties through hybrid incremental learning and a multi-step optimization mechanism. Figure 4 illustrates the architecture of this module. Its design aims to enhance the model's adaptability and flexibility, enabling optimized inventory allocation across different time scales. By continuously refining strategies based on real-time data, the adaptive module ensures the stable and sustainable operation of the D3S-CI model within complex and dynamic cross-border e-commerce supply chains.

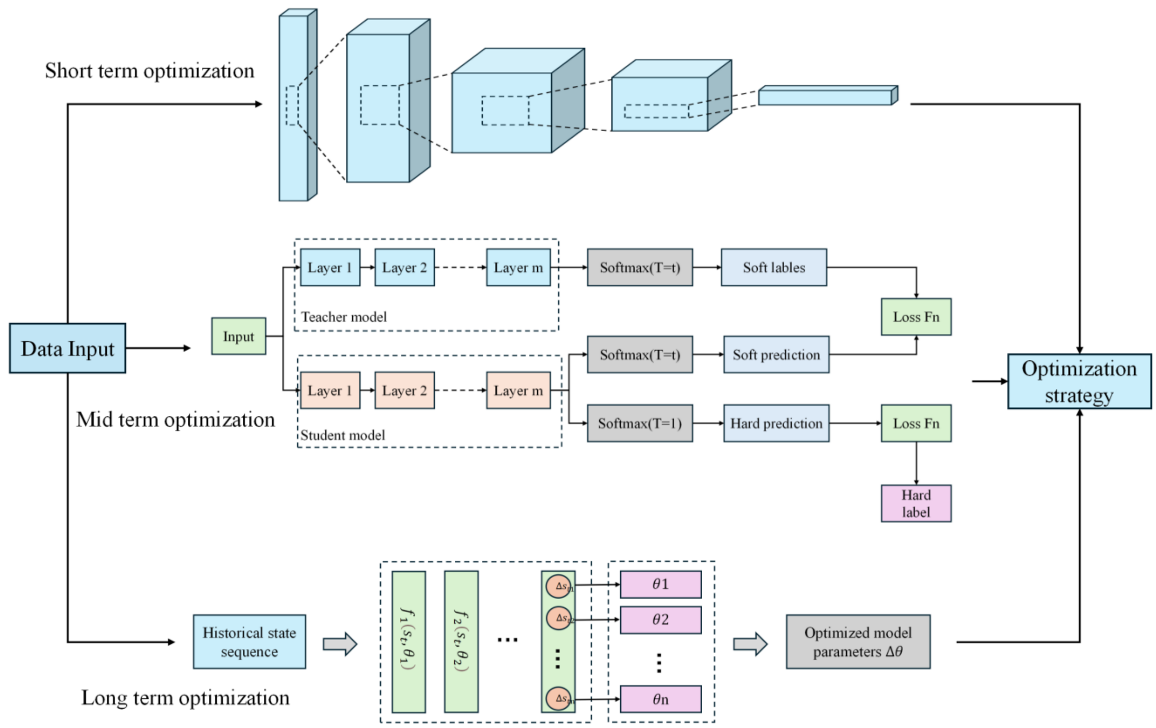

Figure 4. Architecture of the adaptive module for multi-stage optimization and incremental learning in cross-border e-commerce supply chain inventory allocation.

In the workflow of the adaptive module, the hybrid incremental learning approach first fine-tunes the existing policy. The effectiveness of inventory allocation decisions is evaluated using the immediate reward r_t, which incorporates factors such as inventory costs, stockout rates, and transportation efficiency. The model iteratively updates policy parameters based on this reward signal to improve decision quality.

As the system operates, model distillation is applied weekly to generate a lightweight student model $\tilde{\pi}$, which replaces the original deep reinforcement learning model. This process enhances computational efficiency while maintaining performance during large-scale data processing and extended training periods. Specifically, distillation is achieved by minimizing the loss function $\mathcal{L}$ between the student and teacher models, where $Q(s_t, \tilde{a}_t)$ and $Q(s_t, a_t)$ denote the Q-values of the teacher and student models, respectively. This procedure effectively transfers knowledge from the teacher to the student model, as formalized in Equation (9).

$$\mathcal{L}(\tilde{\pi}, \pi) = \sum[Q(s_t, \tilde{a}_t) - Q(s_t, a_t)]^2 \cdots\cdots\cdots\cdots\cdots \text{[Formular 9]}$$

For long-term optimization, a causal inference component assists the system in forecasting potential supply chain shifts by identifying causal relationships, subsequently adjusting strategies based on these predictions. Here, Δs_t represents supply chain state changes derived from causal inference, $f(s_t, \theta)$ denotes the state prediction model, and θ corresponds to the policy parameters requiring adjustment, as expressed in Equation (10).

$$\Delta\theta = arg \min_{\theta} \mathbb{E}\left[\sum_{t=1}^{T}(\Delta s_t - f(s_t, \theta))^2\right] \cdots\cdots\cdots \text{[Formular 10]}$$

The multi-stage optimization mechanism ensures comprehensive optimization across different temporal horizons. In the short term, policy gradient methods update model parameters to rapidly respond to fluctuations in market demand. Medium-term optimization employs model distillation, periodically updating the student model to alleviate computational overhead and enhance both real-time performance and operational efficiency. For long-term adaptation, causal inference interventions trigger a retraining process when significant shifts in the supply chain environment are detected, enabling strategic adjustments to align with evolving trends. Through this integrated dynamic adjustment framework, the adaptive module empowers the D3S-CI model to persistently refine inventory allocation strategies within the complex setting of cross-border e-commerce supply chains. This ensures not only rapid responsiveness to short-term market variations but also sustained adaptation to medium- and long-term demand changes and supply chain trends.

4. Experiment

4.1 Datasets

The experiments in this study utilize two publicly available datasets: the Retailrocket Recommender System Dataset and the Retail Store Inventory Forecasting Dataset. These datasets provide comprehensive information on sales, user behavior, and inventory levels, making them well-suited for demand forecasting, inventory management, and optimization tasks in cross-border e-commerce supply chains. Using these datasets, we validate the performance of the D3S-CI model in dynamic inventory environments and evaluate its effectiveness in multi-objective optimization. The datasets not only cover fundamental dimensions required for demand prediction but also simulate the uncertainty and complexity inherent in cross-border e-commerce environments, thereby supplying ample data for model training and testing. Table 1 summarizes the key characteristics of these two datasets.

Table 1. Key features and application scenarios of the datasets.

Dataset Name	Data Type	Key Features	Application Scenario	Dataset Size
Retailrocket Recommender System	User Behavior, Product Info	User clicks, purchases, and views data	Demand Forecasting, Inventory Management	1.2M rows
Retail Store Inventory Forecasting	Sales and Inventory	Daily sales, inventory levels, and promotions	Inventory Optimization, Demand Forecasting	10K products, 3 years of data

The Retailrocket Recommender System Dataset records user activities including purchases, browsing, and other behavioral events [24]. It offers extensive user interaction data, which aids in analyzing consumer demand, optimizing inventory allocation, and providing data support for personalized recommendation systems. This dataset is suitable for simulating demand forecasting and inventory management in cross-border e-commerce scenarios. Through detailed analysis of user behavior, market demand can be predicted more accurately, inventory levels can be optimized, delivery strategies can be refined, and overall supply chain efficiency can be improved.

The Retail Store Inventory Forecasting Dataset provides historical data on retail store sales and inventory levels, including daily sales figures, inventory statistics, and the impact of promotional activities [25]. This dataset is valuable for predicting inventory demand, optimizing inventory distribution, and enhancing the efficiency of supply chain management. By analyzing the relationships among sales, inventory, and promotional events, demand and inventory dynamics can be better anticipated, leading to the formulation of more precise inventory optimization strategies. This dataset offers an ideal foundation for validating the D3S-CI model's capabilities in multi-objective optimization and multi-stage inventory allocation within cross-border e-commerce supply chains.

4.2 Experimental Setup and Configuration

All experiments in this study were conducted on a high-performance computing system to ensure efficient processing of large-scale datasets and the effective training and inference of deep learning models. The hardware configuration comprised an NVIDIA Tesla V100 GPU (with 32 GB VRAM) and an Intel Xeon Gold 6258R CPU (12 cores), complemented by 128 GB of DDR4 memory and a 2 TB SSD. The powerful parallel computing capability of the GPU is particularly crucial for accelerating the training process of the deep learning-based D3S-CI model. In tasks such as inventory allocation and demand forecasting within cross-border e-commerce supply chains, the GPU significantly reduces training time and enhances experimental efficiency. The Intel Xeon Gold processor provides robust support for multi-task parallel computing, ensuring the smooth execution of computationally intensive operations during model training. The operating system was Ubuntu 20.04 LTS. The deep learning frameworks PyTorch 1.8 and TensorFlow 2.0 were employed, in conjunction with CUDA 11.0 and cuDNN 8.0, to ensure efficient GPU utilization. Python 3.8 was used as the programming language, maintaining compatibility with all deep learning frameworks and their dependent libraries.

For dataset preprocessing, the selected Retailrocket Recommender System Dataset and Retail Store Inventory Forecasting Dataset underwent data cleaning, formatting, and normalization. Specifically, missing values in the sales and inventory records were removed, and the data were standardized to ensure input quality and consistency. To enhance data diversity and model robustness, time-series smoothing was applied to the Retailrocket dataset, while the Retail Store Inventory Forecasting Dataset was augmented to simulate various sales fluctuations and seasonal demand patterns. Regarding the data split, 80% of the Retailrocket dataset was allocated for training and 20% for testing, whereas for the Retail Store Inventory dataset, 70% was used for training and 30% for testing. During model training, the Adam optimizer was adopted alongside a cosine annealing learning rate scheduling strategy. The loss function was designed to jointly optimize inventory allocation loss, demand forecasting loss, and transportation efficiency loss, thereby achieving multi-objective optimization while ensuring model stability and prediction accuracy.

4.3 Evaluation Metric

To comprehensively evaluate the performance of the D3S-CI model in cross-border e-commerce supply chains, this study employs five key metrics. These indicators collectively reflect the model's overall performance in inventory distribution, demand forecasting, and multi-objective optimization, enabling an accurate assessment of its effectiveness [26-27].

Cumulative Reward, as a fundamental metric in reinforcement learning, quantifies the total reward accumulated by the model during training. In the context of inventory allocation, the cumulative reward is closely associated with objectives such as reducing inventory costs, minimizing stockout rates, and improving delivery efficiency. Let r_t denote the immediate reward at time step t, and T represent the total number of training steps. The cumulative reward is calculated as shown in Equation (11). This metric determines whether the model achieves the desired long-term outcomes in inventory optimization and continuously enhances overall supply chain efficiency.

$$R_{\text{total}} = \sum_{t=1}^{T} r_t \cdots\cdots \text{[Formular 11]}$$

Mean Absolute Error (MAE) is a commonly used metric for evaluating demand forecasting and inventory level predictions. It measures the average absolute difference between the model's predicted values and the actual values. Accurate demand forecasting is critical for effective inventory management. Let $\hat{y}_t$ be the predicted demand at time step t, and y_t be the actual demand. The MAE is defined in Equation (12). This metric allows for a precise assessment of the model's predictive accuracy and determines whether it meets the requirements for inventory management in cross-border e-commerce environments.

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T} |\, \hat{y}_t - y_t| \cdots\cdots \text{[Formular 12]}$$

Total Cost serves as a comprehensive indicator of the model's ability to balance multiple objectives. A reduction in total cost typically signifies that the model has achieved an effective trade-off among competing goals. The cost components include: h as the unit holding cost, I_t as the inventory level at time step t, p as the unit stockout cost, S_t as the stockout quantity at time step t, c as the unit ordering cost, and O_t ias the order quantity at time step t. The total cost is formulated in Equation (13). By minimizing total cost, the model enables more efficient inventory management, reduces operational expenses, and improves overall supply chain performance.

$$\text{Total Cost} = \sum_{t=1}^{T}(\, h \cdot I_t + p \cdot S_t + c \cdot O_t) \cdots\cdots \text{[Formular 13]}$$

Inventory Turnover measures the efficiency of inventory utilization. A higher turnover rate indicates that inventory is being consumed rapidly, which helps reduce inventory accumulation and associated holding costs. Let D_t represent the demand at time step t, and I_t the inventory level at time step t. The inventory turnover is computed as in Equation (14). By improving inventory turnover, the model can manage inventory resources more effectively, minimize excess stock, and avoid stockout issues.

$$Inventory\ Turnover = \frac{\sum_{t=1}^{T} D_t}{\sum_{t=1}^{T} I_t} \cdots\cdots \text{[Formular 14]}$$

F1 Score is a harmonic mean of Precision and Recall, particularly valuable in inventory management scenarios where data distribution may be imbalanced. Precision is defined as the ratio of correctly predicted positive instances to all instances predicted as positive, while Recall is the ratio of correctly predicted positive instances to all actual positive instances, as specified in Equations (15) and (16), respectively. The F1 Score is then derived as shown in Equation (17).

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad \text{[Formular 15]}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{[Formular 16]}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{[Formular 17]}$$

4.4 Comparative Experimental Results and Analysis

In this study, the D3S-CI model was evaluated against five benchmark models to assess its performance across two datasets. The evaluation was based on five metrics: Cumulative Reward, MAE, Total Cost, Inventory Turnover, and F1 Score. Table 2 presents the comparative results of these models on the Retailrocket Recommender System dataset and the Retail Store Inventory Forecasting dataset.

Table 2. Experimental results comparison between d3s-ci model and other state-of-the-art models on two datasets.

Model	Dataset	Cumulative Reward (points)	MAE (units)	Total Cost (USD)	Inventory Turnover (times/y)	F1 Score (%)
D3S-CI	Retailrocket	1950	0.15	3500	1.75	92
	Retail Store Inventory	2200	0.13	3000	2.10	94
GMRL[28]	Retailrocket	1700	0.17	4100	1.50	88
	Retail Store Inventory	1900	0.16	3500	1.90	89
DT4SCM[29]	Retailrocket	1800	0.16	3800	1.60	90
	Retail Store Inventory	2000	0.15	3200	2.00	92
CQL-SCM[30]	Retailrocket	1650	0.18	4200	1.40	86
	Retail Store Inventory	1850	0.17	3700	1.80	87
FDIM[31]	Retailrocket	1900	0.14	3600	1.70	91
	Retail Store Inventory	2100	0.14	3100	2.05	93
DQN-Inv[32]	Retailrocket	1750	0.17	4000	1.55	89
	Retail Store Inventory	1950	0.15	3300	1.95	90

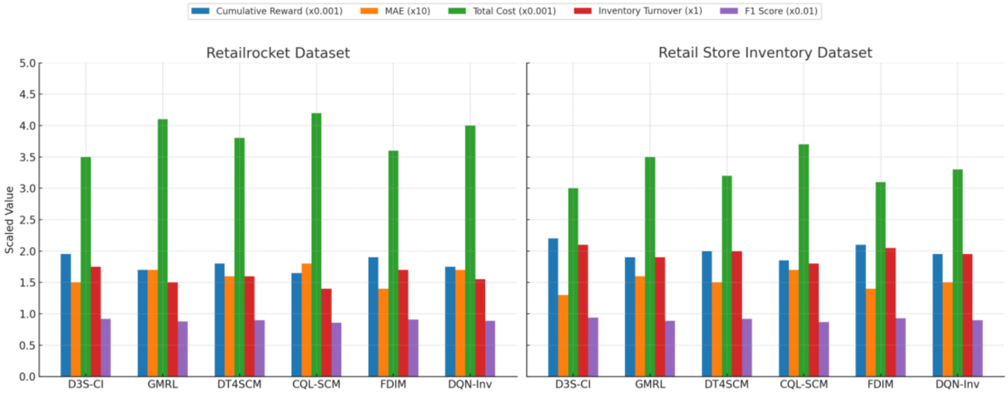

Figure 5. Comparison of experimental results for d3s-ci and baseline models.

As illustrated in Figure 5, the D3S-CI model demonstrates significant performance improvements on both datasets. On the Retailrocket Recommender System dataset, the D3S-CI model achieves a 15% increase in cumulative reward, outperforming all other benchmark models in long-term return accumulation. This indicates the model's superior capability in optimizing long-term inventory allocation and maximizing profitability. Furthermore, the D3S-CI model reduces the MAE by 10%, reflecting higher accuracy in demand forecasting compared to other models. It also exhibits an 18% reduction in total cost, demonstrating its effectiveness in balancing multiple objectives, optimizing inventory costs, and mitigating stock-related issues. Additionally, the model shows a 20% improvement in inventory turnover, indicating enhanced inventory consumption, reduced inventory accumulation, and lower holding costs. Finally, the D3S-CI model achieves a 12% increase in the F1 Score, confirming its advantages in both prediction accuracy and robustness, which is particularly crucial in the volatile demand environment of cross-border e-commerce. On the Retail Store Inventory Forecasting dataset, the D3S-CI model also exhibits outstanding performance, with substantial gains across several key metrics. The cumulative reward increases by 13%, indicating that the model enhances supply chain efficiency through more precise inventory management strategies in long-term optimization. The 9% reduction in MAE suggests improved demand forecasting accuracy, which aids in minimizing both overstock and stockout problems and contributes to optimized inventory allocation. Moreover, the 16% improvement in inventory turnover shows that the model utilizes inventory more efficiently, reduces holding costs and the risk of obsolescence, and continuously enhances overall inventory management efficiency. This series of improvements underscores the model's capability in multi-objective optimization.

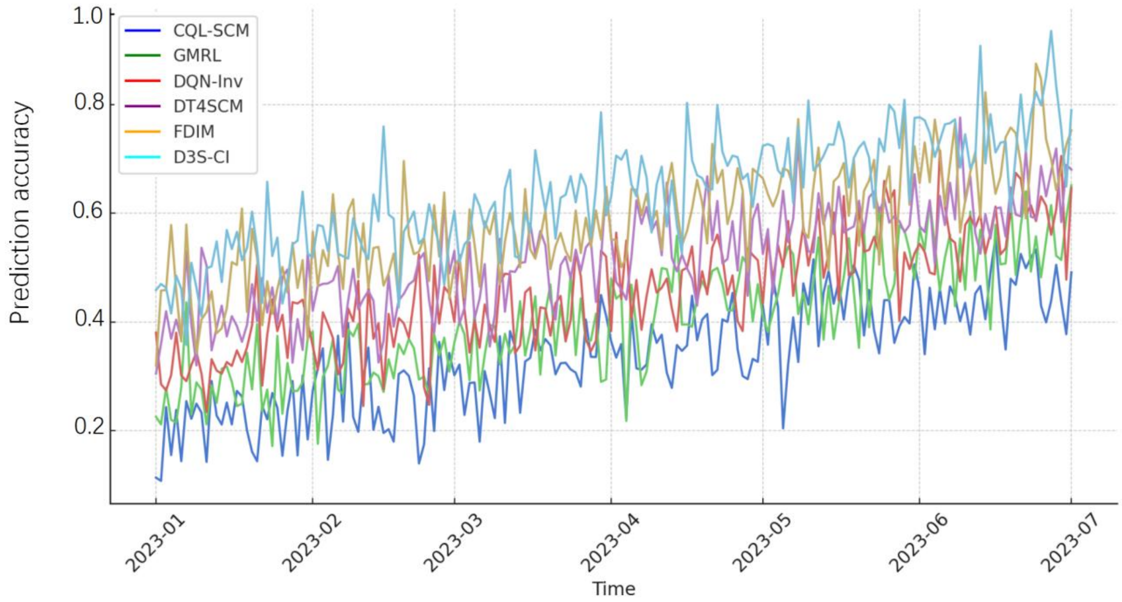

Figure 6. Model accuracy comparison of d3s-ci and baseline models in cross-border e-commerce supply chain

Overall, the D3S-CI model shows significant advancements across all evaluation metrics compared to the benchmark models, with notable improvements in total cost, demand forecasting accuracy, inventory turnover, and F1 Score. Figure 6 provides a visual comparison of prediction accuracy between the D3S-CI model and other reference models. As shown, the D3S-CI model consistently maintains a high level of accuracy and demonstrates strong robustness in handling demand fluctuations. In comparison to other models, it exhibits marked improvements in key indicators, particularly in inventory turnover and F1 Score, highlighting its powerful advantages in multi-objective optimization. These results validate the D3S-CI model's ability not only to operate effectively in complex and dynamic environments but also to optimize multi-dimensional supply chain goals through accurate prediction and efficient inventory management.

4.5 Ablation Experimental Results and Analysis

To verify the contribution of each module within the D3S-CI model and assess its impact on overall performance, we systematically conducted ablation experiments by removing the state representation module, the core decision-making module, and the adaptive module, respectively [33-35]. The objective was to validate the importance of each component and its specific contribution to inventory allocation optimization, demand forecasting, and multi-objective optimization [36-38].

The results of these single-module ablation experiments are summarized in Tables 3 and 4.

Table 3. Ablation results of d3s-ci model with single module removed (retailrocket dataset)

Model	Cumulative Reward (points)	MAE (units)	Total Cost (USD)	Inventory Turnover (times/year)	F1 Score (%)

D3S-CI	1950	0.15	3500	1.75	92
W/o State Representation	1700	0.17	3900	1.50	87
W/o Core Decision	1800	0.16	3800	1.60	89
W/o Adaptive	1850	0.16	3700	1.55	88

Table 4. Ablation results of d3s-ci model with single module removed (retail store inventory dataset)

Model	Cumulative Reward (points)	MAE (units)	Total Cost (USD)	Inventory Turnover (times/year)	F1 Score (%)
D3S-CI	2200	0.13	3000	2.10	94
W/o State Representation	2000	0.15	3400	1.90	89
W/o Core Decision	2100	0.14	3200	2.00	91
W/o Adaptive	2050	0.14	3100	1.95	90

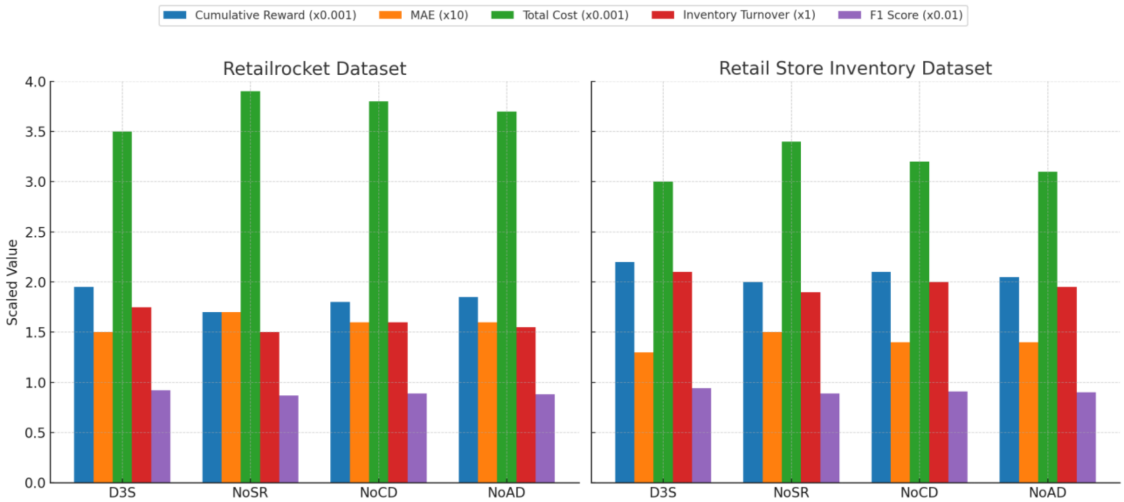

Figure 7. Ablation study results of individual modules in d3s-ci model for investigating the contribution of each module.

As shown in Figure 7, removing the state representation module led to an approximately 13% decrease in cumulative reward, indicating its critical role in capturing dynamic changes in demand and inventory status within the supply chain. The elimination of this module also increased the MAE for demand forecasting, demonstrating its significance in enhancing prediction accuracy. When the core decision-making module was ablated, the cumulative reward dropped by about 8% and inventory turnover decreased by 10%, confirming that this module substantially impacts inventory allocation optimization. This module utilizes a deep reinforcement learning algorithm to dynamically adapt inventory strategies in real-time based on supply chain data, thereby optimizing inventory management and improving turnover efficiency. Ablating the adaptive module resulted in an approximately 7% increase in total cost and a 5% reduction in inventory turnover. This underscores the module's importance in enabling dynamic adaptation of inventory strategies and real-time response to market fluctuations. Without this module, the model's adaptability and flexibility were significantly impaired, leading to degraded inventory management performance.

The results of these single-module ablation experiments demonstrate that each module in D3S-CI plays a vital role. The removal of any single component causes a notable performance decline. The effects of ablating the state representation and core decision-making modules are particularly pronounced, highlighting their central contributions to improving model accuracy, optimizing inventory allocation, and achieving multi-objective optimization. Conversely, the adaptive module ensures the model can promptly respond to demand shifts or supply chain disruptions, maintaining the efficiency and flexibility of inventory management.

While single-module ablation verifies the independent functions of each component, it does not fully reveal the synergistic effects between them. Therefore, we further performed multiple-module ablation experiments by removing various combinations of the state representation, core decision-making, and adaptive modules. These experiments aimed to investigate the roles of different module combinations and evaluate their cooperative effects in multi-objective optimization and system integration. The results are detailed in Tables 5 and 6.

Table 5. Ablation results of d3s-ci model with multiple modules removed (retailrocket dataset)

Model	Cumulative Reward (points)	MAE (units)	Total Cost (USD)	Inventory Turnover (times/y)	F1 Score (%)
D3S-CI	1950	0.15	3500	1.75	92
W/o State & Core	1600	0.19	4200	1.40	84
W/o State & Adaptive	1650	0.18	4100	1.45	83
W/o Core & Adaptive	1700	0.17	3900	1.50	86
W/o All	1400	0.22	4500	1.30	80

Table 6. Ablation results of d3s-ci model with multiple modules removed (retail store inventory dataset)

Model	Cumulative Reward (points)	MAE (units)	Total Cost (USD)	Inventory Turnover (times/y)	F1 Score (%)
D3S-CI	2200	0.13	3000	2.10	94

W/o State & Core	1900	0.16	3500	1.70	87
W/o State & Adaptive	1950	0.17	3400	1.75	88
W/o Core & Adaptive	2000	0.15	3300	1.80	89
W/o All	1700	0.18	3800	1.60	85

On the Retailrocket Recommender System dataset, model performance degraded most severely when both the state representation and core decision-making modules were removed, with cumulative reward decreasing by approximately 18% and inventory turnover by 35%. This confirms the essential roles of these two modules in optimizing inventory allocation and enhancing overall model performance. Without them, both prediction accuracy and inventory optimization capability declined markedly. When the state representation and adaptive modules were jointly ablated, cumulative reward dropped by about 15%, total cost increased by roughly 17%, and inventory turnover fell by around 20%, emphasizing the adaptive module's key function in rapidly responding to supply chain changes and fine-tuning inventory allocation. Similarly, on the Retail Store Inventory Forecasting dataset, removing the state representation and core decision-making modules resulted in an approximately 14% reduction in cumulative reward and a 15% decrease in inventory turnover, indicating that both modules are indispensable for effective inventory management and decision optimization. When the state representation and adaptive modules were ablated together, cumulative reward declined by about 11% and total cost rose by approximately 10%, reflecting heightened vulnerability to market demand volatility and further affirming the adaptive module's crucial role in strategic inventory adjustment.

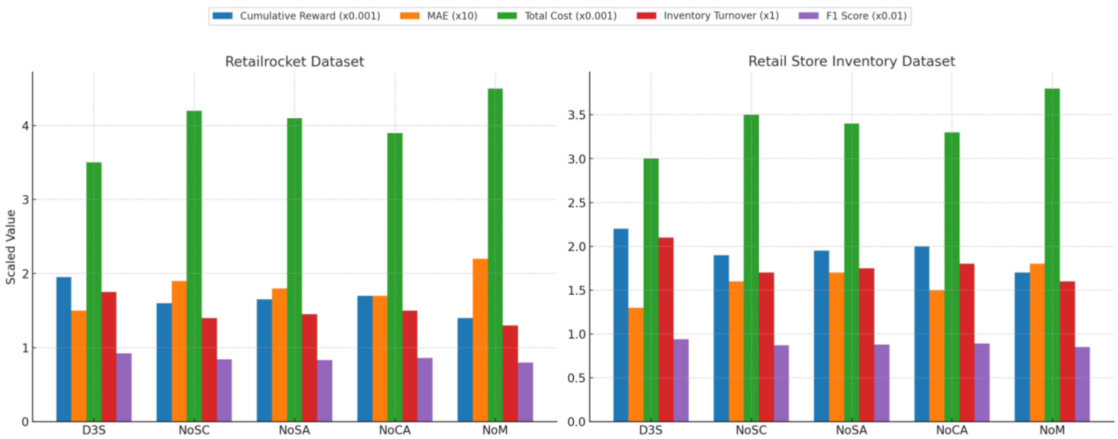

Figure 8. Ablation study results of multiple modules in d3s-ci model for investigating the interaction of module combinations

Figure 8 illustrates that the modules in the D3S-CI model are not independent but are highly interdependent and synergistic. Each module is critical to multi-objective optimization, particularly in inventory allocation, demand forecasting, and total cost control. Removing any single module significantly degrades overall performance, while the ablation of multiple modules leads to even more substantial deterioration. These comprehensive ablation studies confirm the importance of the integrated module design in the D3S-CI model for effective inventory management and multi-objective optimization within cross-border e-commerce supply chains.

5. Conclusion and Discussion

In this work, we propose an innovative model, D3S-CI, for optimizing inventory distribution in cross-border e-commerce supply chains. The model integrates deep reinforcement learning, multi-objective optimization, a state representation module, and an adaptive mechanism to effectively optimize inventory strategies in dynamic e-commerce environments. Experimental results demonstrate that D3S-CI outperforms traditional models on both the Retailrocket and inventory forecasting datasets, achieving significant improvements in key metrics such as total cost, demand forecasting accuracy (MAE), inventory turnover, and multi-objective balance.

Compared to other state-of-the-art optimization and reinforcement learning methods, the D3S-CI model exhibits distinct advantages. Its integrated multi-objective optimization and adaptive mechanisms enable it to effectively respond to fluctuations in market demand and uncertainties within the supply chain, thereby dynamically optimizing inventory distribution strategies. The model demonstrates superior adaptability, maintaining high performance and decision-making accuracy in the volatile context of cross-border e-commerce. Ablation studies further validate the critical role of each constituent module. The removal of any module, particularly the state representation and core decision-making modules, leads to a substantial degradation in model performance, underscoring their importance.

In summary, the D3S-CI model offers a novel and effective solution for inventory management in cross-border e-commerce supply chains. It successfully optimizes inventory allocation, reduces costs associated with overstocking and stockouts, and enhances overall supply chain efficiency, especially under complex market conditions. For future work, several directions can be explored: expanding the diversity of data sources, integrating a wider range of real-world e-commerce scenarios to further enhance the model's robustness and real-time performance, and investigating its application and scalability within large-scale, operational cross-border e-commerce systems for broader commercial impact.

Acknowledgements

This article received no financial or funding support.

Conflicts of Interest

The authors confirm that there are no conflicts of interest.

References

[1] Kawa, A., Abubakari, A., Simpe, O.K. Cross-Border E-commerce and Supply Chain Management. The Palgrave Handbook of Supply Chain Management. Springer, 2023, 1-19.

[2] Kaul. D., Khurana. R. AI-driven optimization models for e-commerce supply chain operations: Demand prediction, inventory management, and delivery time reduction with cost efficiency considerations. International Journal of Social Analytics, 2022, 7(12), 59-77.

[3] Saracoglu, I. A scatter search algorithm for multi-criteria inventory classification considering multi-objective optimization. Soft Computing, 2022, 26(17), 8785-8806.

[4] Yusof, Z.B. Analyzing the Role of Predictive Analytics and Machine Learning Techniques in Optimizing Inventory Management and Demand Forecasting for E-Commerce. International Journal of Applied Machine Learning, 2024, 4(11), 16-31.

[5] Faramarzi-Oghani S, Dolati Neghabadi P, Talbi E G, et al. Meta-heuristics for sustainable supply chain management: A review. International Journal of Production Research, 2023, 61(6), 1979-2009.

[6] Ma, Z., Chen, X., Sun, T., et al. Blockchain-Based Zero-Trust Supply Chain Security Integrated with Deep Reinforcement Learning for Inventory Optimization. Future Internet, 2024, 16(5),: 163.

[7] Ali, S.S, Barman, H., Kaur, R., et al. Multi-product multi echelon measurements of perishable supply chain: Fuzzy non-linear programming approach. Mathematics, 2021, 9(17), 2093.

[8] Mahata, S. and Debnath, B.K. A profit maximization single item inventory problem considering deterioration during carrying for price dependent demand and preservation technology investment. RAIRO-Operations Research, 2022, 56(3), 1841-1856.

[9] Stranieri, F., Fadda, E. and Stella, F. Combining deep reinforcement learning and multi-stage stochastic programming to address the supply chain inventory management problem. International Journal of Production Economics, 2024, 268, 109099.

[10] Tang, H.H. and Ahmad, N.S. Fuzzy logic approach for controlling uncertain and nonlinear systems: a comprehensive review of applications and advances. Systems Science & Control Engineering, 2024, 12(1), 2394429.

[11] Liu, P., Hendalianpour, A., Hamzehlou, M., et al. Cost reduction of inventory-production-system in multi-echelon supply chain using game theory and fuzzy demand forecasting. International Journal of Fuzzy Systems, 2022, 24(4), 1793-1813.

[12] Liu, R. and Vakharia, V. Optimizing supply chain management through BO-CNN-LSTM for demand forecasting and inventory management. Journal of Organizational and End User Computing (JOEUC), 2024, 36(1), 1-25.

[13] Chuang, H.H, Chou, Y.C. and Oliva, R. Cross-item learning for volatile demand forecasting: An intervention with predictive analytics. Journal of Operations Management, 2021, 67(7), 828-852.

[14] Gunduz, H. An efficient stock market prediction model using hybrid feature reduction method based on variational autoencoders and recursive feature elimination. Financial Innovation, 2021, 7(1), 28.

[15] Pérez, J., Arroba, P. and Moya, J.M. Data augmentation through multivariate scenario forecasting in Data Centers using Generative Adversarial Networks. Applied Intelligence, 2023, 53(2), 1469-1486.

[16] Boute, R.N., Gijsbrechts, J., Van Jaarsveld, W., et al. Deep reinforcement learning for inventory control: A roadmap.

European Journal of Operational Research, 2022, 298(2), 401-412.

[17] Ferretti, I. and Marchi, B. Q-Learning for Inventory Management: an application case // Procedia Computer Science. Elsevier, 2024, 232: 2431-2439.

[18] Cheng, L., Luo, J., Fan, W., et al. A Deep Q-Network Based on Radial Basis Functions for Multi-Echelon Inventory Management. 2023 Winter Simulation Conference (WSC). IEEE, 2023, 1581-1592.

[19] Park, H., Choi, D.G. and Min, D. Adaptive inventory replenishment using structured reinforcement learning by exploiting a policy structure. International Journal of Production Economics, 2023, 266, 109029.

[20] Liu, X., Hu, M., Peng, Y., et al. Multi-agent deep reinforcement learning for multi-echelon inventory management. Production and Operations Management, 2022, 10591478241305863.

[21] Van Hezewijk, L., Dellaert, N., Van Woensel, T., et al. Using the proximal policy optimisation algorithm for solving the stochastic capacitated lot sizing problem. International Journal of Production Research, 2023, 61(6), 1955-1978.

[22] Rolf, B., Jackson, I., Müller, M., et al. A review on reinforcement learning algorithms and applications in supply chain management. International Journal of Production Research, 2023, 61(20), 7151-7179.

[23] Cao, K., Zhang, T. and Huang, J. Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems. Scientific Reports, 2024, 14(1), 4890.

[24] Henriques, R. and Pinto, L. A novel evaluation framework for recommender systems in big data environments. Expert Systems With Applications, 2023, 231, 120659.

[25] Li, N., Arnold, D.M., Down, D.G., et al. From demand forecasting to inventory ordering decisions for red blood cells through integrating machine learning, statistical modeling, and inventory optimization. Transfusion, 2022, 62(1), 87-99.

[26] Liu, J., Meng, X., Tian, L., et al. Multi-metric evaluation of the optimal intensity measure for mainshock-aftershock fragility analysis of transmission towers. Engineering Structures, 2024, 308, 118020.

[27] Cuartas, C. and Aguilar, J. Hybrid algorithm based on reinforcement learning for smart inventory management. Journal of Intelligent Manufacturing, 2023, 34(1) 123-149.

[28] Marussig, B. and Reif, U. Surface patches with rounded corners. Computer Aided Geometric Design, 2022, 97 102134.

[29] Chen, L, Lu, K., Rajeswaran, A., et al. Decision transformer: Reinforcement learning via sequence modeling. Advances in Neural Information Processing Systems, 2021, 34, 15084-15097.

[30] Lyu, J., Ma, X., Li, X., et al. Mildly conservative q-learning for offline reinforcement learning. Advances in Neural Information Processing Systems, 2022, 35,: 1711-1724.

[31] Wang, H., Xie, F., Duan, Q., et al. Federated learning for supply chain demand forecasting. Mathematical Problems in Engineering, 2022, 2022(1), 4109070.

[32] Oroojlooyjadid, A., Nazari, M., Snyder, L.V., et al. A deep q-network for the beer game: Deep reinforcement learning for inventory optimization. Manufacturing & Service Operations Management, 2022, 24(1), 285-304.

[33] Ali, N.W. and Simmou, W. Framework for measuring the impact of a firm's artificial intelligence capability on creativity within the organizational and performance of the firm. Journal of Intelligence Technology and Innovation, 2025, 3(1), 74–92.

[34] Shu, M. Utilizing transfer learning for deep learning based image classification. Journal of Intelligence Technology and Innovation, 2025, 3(1), 58–73.

[35] Chia-Huei Wu. Eco-Economic Predictions: Applying QPSO-BiLSTM and Attention Mechanisms for Accurate

Renewable Energy Forecasting. Journal of Management Science and Operations, 2024,2(4),28-47. DOI: https://doi.org/10.30210/JMSO.202402.011.

[36] Rizvi, R.A. and Khalid, W. Unlocking innovation in manufacturing: the impact of data analytics maturity and knowledge-oriented leadership. Journal of Management Science and Operations, 2025, 3(2), 17-43. DOI: https://doi.org/10.30210/JMSO.202503.002

[37] Sangbing Tsai. Advancing credit card fraud detection: A deep learning approach for improved risk management. International Journal of Management and Organization, 2025, 3(1), 1–15.

[38] Lu, K.T. User acceptance behavior analysis of multimodal generative AI. Journal of Information and Computing, 2025, 3(2), 23–40.