# Supply Chain Logistics Demand Modeling and Abnormal Warning Based on Autoformer

Chuanchuan Teng, and Xizhuo Han*

***Abstract*—Current methods have limited capabilities in accurately modeling and dynamically warning of anomalies in port supply chain logistics (SCLs) scenarios where long-term trends in logistics demand (LD) coexist with short-term disturbances. To improve the accuracy of LD forecasting and anomaly warning (AW), and enhance the stability and responsiveness of the supply chain (SC), this paper constructed a multi-scale demand forecasting and AW model based on Autoformer. First, multi-source logistics data was integrated and combined with the Autoformer trend-disturbance decomposition mechanism to achieve structured modeling of non-stationary demand sequences. Then, a multi-scale attention path was designed to extract periodic information at the daily, monthly, and quarterly levels, respectively, to enhance the model's perception of multi-frequency features. Finally, a residual-driven anomaly scoring function and a dynamic threshold strategy were combined to construct a multi-level AW mechanism to achieve accurate identification of logistics risk events of different intensities. Experimental results show that in LD modeling, the Autoformer achieves a final weighted average percentage wrror (WAPE) and root mean square error (RMSE) of 0.105, and a matching degree of over 0.900 for the changing trend of supply chain logistics demand (SCLD) for a 200 twenty-feet equivalent unit (TEU) SC. In terms of AW, the Autoformer achieves average false negative rate (FNR) and false positive rate (FPR) of 12.0% and 7.3%, respectively, with an average overall delay of 1.3 hours. The weighted Kappa coefficient for consistency in AW levels is 0.844. The results show that the proposed Autoformer model, on the port logistics dataset used, demonstrates certain improvements in demand forecasting and anomaly detection compared to the selected baseline model, providing a feasible solution reference for forecasting and early warning tasks in intelligent supply chain management.**

***Index Terms*—Supply Chain Logistics, Abnormal Warning, Autoformer Model, Multi-Scale Attention Mechanism, Trend Perturbation Decomposition**

## I. INTRODUCTION

WITH the deep penetration of information technologies such as the Internet of Things, 5G communication and cloud computing into the logistics field, the modern supply chain has evolved into a highly information-based and interconnected complex system. Real-time information interaction between logistics nodes and the fusion processing of multi-source heterogeneous data have become the key foundation for supporting accurate decision-making and agile response in the supply chain. Against this background, port logistics demand (PLD) exhibits typical characteristics of multivariability, high dynamism and multi-scale, with complex data dimensions and intertwined influencing factors [1], [2]. Introducing cutting-edge deep learning methods suitable for high-dimensional time series data processing into supply chain logistics modeling is not only a natural extension of the method, but also an inherent requirement for the digital transformation and upgrading of the industry. Existing risk warning models struggle to accurately capture the correlations between key indicators such as cargo throughput, ship arrival frequency, and warehouse turnover [3], [4], [5]. Faced with both long-term trend fluctuations and short-term sudden fluctuations in PLD, these models struggle to address their non-stationarity, the coexistence of long-term and short-term trends, and sudden disruptions. This can lead to inaccurate forecasts and delayed AWs, severely restricting the rational allocation of port logistics resources and the timeliness of emergency responses.

This paper combines the Autoformer model [6]with the structural characteristics of SC port logistics scenarios to construct an enhanced prediction and anomaly detection (AD) framework. By constructing an adaptive trend decomposition module and a multi-scale attention mechanism (MSAM), this paper achieves a high-precision forecasting and dynamic AW of PLD, which has important theoretical value and practical significance for improving the logistics response efficiency and risk prevention and control level of ports.

The main contributions of this paper are reflected in the integration and scenario-specific improvements of existing time-series modeling components tailored to the characteristics of SCLs scenarios:

1) A "trend-disturbance dual-path adaptive decomposition architecture" for the non-stationary demands of port logistics is proposed. This architecture improves the residual decomposition problem of the original Autoformer under strong sudden disturbances by using a learnable sliding average window and residual paths.

2) A nested multi-scale attention fusion mechanism is designed. Through learnable scale weights, it achieves adaptive extraction and fusion of multi-level periodic features at the daily, monthly, and quarterly levels, overcoming the limitations of fixed scale division.

School of Economics and Management, Cangzhou College of Transportation, Cangzhou City, Hebei Province, China
* Corresponding Author, e-mail: hanxizhuo0313@163.com

DOI: 10.36244/ICJ.2026.2.5

3) A dynamic residual scoring and multi-level threshold early warning system is constructed. Combining local statistics of the sliding window and duration judgment, it achieves graded and sensitive responses to anomalies of different degrees in logistics scenarios.

## II. RELATED WORK

In recent years, the informatization and intelligent transformation of SC logistics systems has become a consensus in academia and industry. Real-time monitoring of logistics processes, multi-source data fusion and intelligent decision support are all inseparable from the support of information and communication technology (ICT) and artificial intelligence methods [7]. In particular, with the implementation of concepts such as industrial internet and digital twins, logistics systems have gradually built a two-way mapping of "physical process-information space", which provides rich application scenarios and data foundation for data-driven methods such as deep learning [8], [9]. Some scholars used statistical and neural network models to model SCLs demand, namely, the combination of long short-term memory (LSTM) and auto-regressive integrated moving average (ARIMA) models, as well as neural network and grey prediction models, to improve forecasting accuracy [10]. Zhu Junlin built a data-driven prediction framework by enhancing the LSTM structure, which not only significantly reduced the RMSE but also achieved a 95% true positive rate in AD, demonstrating strong responsiveness [11]. In addition, unsupervised methods such as isolation forests have also been used for AD, especially in practical scenarios where labels are scarce [12]. Yipeng Chen focused on the changes in prediction accuracy at different time granularities and proposed a multi-granularity joint optimization framework to improve resource coordination issues at all levels of the SC [13]. Xie Luqiang proposed a warning model that combines rough sets with radial basis function neural networks to address the hierarchical complexity of cruise ship construction logistics needs, improving the prediction efficiency and emergency response capabilities in specific scenarios [14]. These studies have provided a wealth of methodological support for SCLs' demand forecasting and anomaly identification, but most methods have the limitation of insufficient ability to model long-term dependent features.

To further improve the ability to model long-term dependent features, some researchers have begun exploring modeling methods based on Autoformers, leveraging their built-in autocorrelation mechanisms and decomposition structures to enhance their ability to capture periodicity and trends. Ma Xiangkai, addressing the limitations of traditional Autoformers in modeling multi-scale features, proposed applying multi-scale convolution operations and a temporal encoding mechanism, and constructed an Autoformer model that significantly improved RMSE performance in multi-domain prediction tasks [15]. Cao Danyang designed a more efficient Autoformer structure and added a position embedding mechanism to improve the model training speed and generalization ability, achieving improvements in multiple evaluation indicators [16]. Based on Autoformer, these studies have preliminarily verified its potential in processing complex SC time series data. However, existing multi-scale Transformer methods mostly employ fixed or predefined scale divisions, making it difficult to adaptively capture the complex patterns of interwoven daily, monthly, and quarterly cycles in port logistics. There is also limited research on decoupling the differentiated impacts of trends and disturbances at different scales. While current decomposition-based Transformer models can separate trend and seasonal terms, in port logistics scenarios, short-term sudden disturbances are easily smoothed out or remain in the trend term, leading to decreased anomaly detection sensitivity. Furthermore, most residual-driven anomaly detection methods rely on global or static thresholds, making it difficult to adapt to changes in local statistical characteristics under dynamic fluctuations in logistics demand, and lacking the ability to provide tiered early warnings for anomalies of varying severity.

## III. SUPPLY CHAIN LOGISTICS DEMAND MODELING AND ABNORMAL WARNING BASED ON AUTOFORMER

### A. Supply Chain Multi-Source Heterogeneous Data Fusion

To address the combined impact of historical order volume, inventory turnover, shipping time, and heterogeneous market environments on changing demand in SCLs scenarios, this paper constructs a unified input feature system. First, a multi-source, heterogeneous time series data fusion framework for port logistics scenarios is designed. Through data fusion and standardization, an input dataset with a unified time series structure is constructed.

The multi-source data used in this paper primarily encompasses four dimensions: historical order data, inventory information, shipping time, and market environment indicators. Table I illustrates the original data format.

TABLE I
DATA CHARACTERISTICS OF DIFFERENT DIMENSIONS

| Data type | Feature Name | Time granularity | Unit |
|---|---|---|---|
| Historical order data | Daily order quantity | day | TEU |
| Inventory data | Current inventory level | day | ton |
| Shipping time | Average shipping time | hour | hour |
| Market environment | Consumer Price Index (CPI) | month/day | Index/Boolean |
| | Purchasing Managers´Index (PMI) | | |
| | Holiday signs | | |

To achieve a unified input format, this paper resamples all data at a daily granularity, using nearest neighbor filling for monthly data, daily average aggregation for hourly data, and Boolean flags for holidays (0 for non-holidays, 1 for holidays). To address irregular data sampling, transmission delays, and missing values, this paper uses time index alignment and linear interpolation to align and fill missing values. Using the order data timeline as the primary index, outer joins are used for the remaining dimensions to generate a unified date-aligned data table. Linear interpolation is used for continuous variables:

$$\hat{x}_t = x_t + \frac{(x_{t_2} - x_{t_1})(t - t_1)}{t_2 - t_1},\ t \in (t_1, t_2) \quad (1)$$

$x_{t_1}$ and $x_{t_2}$ are the nearest non-missing values before and after the missing point. $t_1$ and $t_2$ are the nearest valid time points before and after the missing value, respectively. Forward filling is used for discrete or Boolean variables. Due to the significant differences in the dimensions and numerical distribution of each variable, directly inputting the model leads to problems such as uneven feature response and unstable gradients. This paper uses Z-score normalization for all continuous features:

$$z_{i,t} = \frac{x_{i,t} - \mu_i}{\sigma_i} \quad (2)$$

Here, $\mu_i$ and $\sigma_i$ are the mean and standard deviation of the feature in the training set, and $x_{i,t}$ signifies the original value of the i-th feature at time $t$. After normalization, construct the sliding window input sequence $X_t \in \mathrm{R}^{\omega \times d}$ [17], [18]:

$$X_t = [x_{t-\omega+1}, x_{t-\omega+2}, \cdots, x_t]^{\mathrm{T}} \quad (3)$$

The final result is a fixed-length sliding sequence input set $\{X_t, y_t\}$, where $y_t$ is the corresponding target prediction value. To enhance the model's ability to perceive temporal position information and periodic fluctuations, timestamp features are added to the final input, including the week number $Day\ of\ week_t \in \{1,2,3,4,5,6,7\}$, $Month_t \in \{1,2,\cdots,12\}$, and holiday indicator variable $Holiday_t \in \{0,1\}$ at the current time step $t$. All processed timestamp variables are concatenated column by column to obtain the structured time feature matrix $T_t \in R^m$ corresponding to each time step, where $m$ is the total dimension after encoding. To improve the model's ability to perceive temporal position information, a continuous position encoding based on sine and cosine functions is applied [19], [20]:

$$PE_{(t,2i)} = sin(\frac{t}{10000^{2i/d}}) \quad (4)$$

$$PE_{(t,2i+1)} = cos(\frac{t}{10000^{2i/d}}) \quad (5)$$

Finally, this paper concatenates the encoded structured timestamp feature $T_t$ and the continuous position embedding $PE_t$ in the feature dimension to form an enhanced temporal structure feature vector $\tilde{T}_t \in \mathrm{R}^{m+d}$. This vector is input into the model together with the standardized main sequence data.

### B. Supply Chain Trend-Disturbance Adaptive Decomposition

SCLs demand that time series contain long-term trends, cyclical fluctuations, and sudden disturbances [21], [22]. Unified encoding methods are prone to reducing the model's time series prediction accuracy and anomaly identification capabilities due to dynamic feature aliasing [23], [24]. In order to achieve effective analysis and prediction of non-stationary supply chain circulation demand, this paper uses the Autoformer model for adaptive trend decomposition. The input multivariate time series is structurally decomposed to construct trend path and interference path structures to achieve feature transfer processing, as shown in Figure 1.

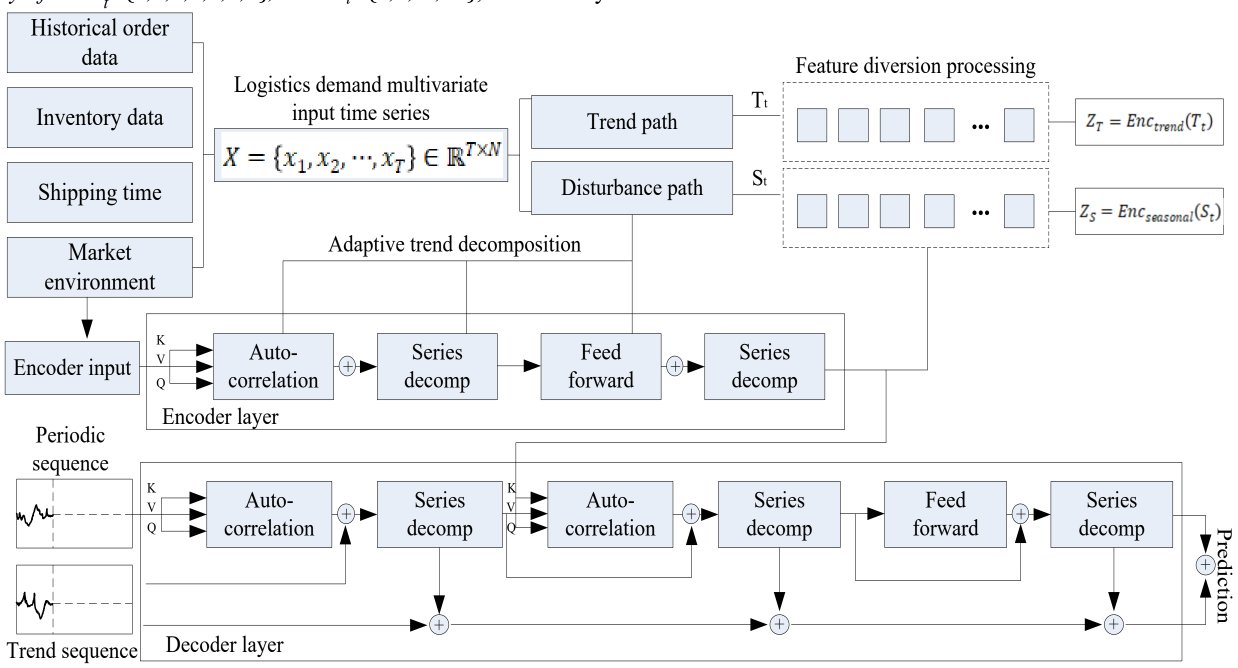

Fig. 1. Adaptive Trend Decomposition and Feature Diversion Modeling

Considering a multivariate input time series $X = \{x_1, x_2, \cdots, x_T\} \in R^{T \times N}$, and the decomposition goal is:

$$X_t = T_t + S_t \quad (6)$$

$T_t$ represents the trend term, which represents the low-frequency trend structure that changes over time; $S_t$ represents the residual term, which contains local perturbations and high-frequency variation components. In order to achieve a learnable decomposition operation, this paper uses a sliding average filter as an approximate low-pass filter. Given a multivariate time series $X=\{x_1, x_2, \cdots, x_T\} \in R^{T \times N}$, and assuming the half-length of the sliding window is $l$ ($l$ is a positive integer), the trend term $T_t$ at time point t is calculated as follows:

$$T_t = \frac{1}{2l+1} \sum_{i=-l}^{l} x_{t+i}, \ t \in [l+1, T-l] \tag{7}$$

In formula, $x_{t+i}$ represents the observation vector at time point $t+i$. The remaining part constitutes the residual term $S_t$:

$$S_t = X_t - T_t \tag{8}$$

The sliding average window is constructed as a learnable parameter, and the window length and weight are automatically optimized through the training process to achieve dynamic adaptability to different data sources. In the encoder part, two path diversion attention modules are designed, namely the trend path and the perturbation path, to perform structural modeling on $T_t$ and $S_t$, respectively, to implement a dynamic feature diversion strategy. The trend term has strong long-term dependence and slowly changing characteristics, and it is suitable to use an attention mechanism with a larger receptive field to extract global context information. An autocorrelation-based structural attention extractor is used to replace the standard dot product attention calculation with time-delayed correlation estimation to improve efficiency and long-term modeling capabilities:

$$AutoCoor(Q,K) = \frac{1}{T} \sum_{\tau=0}^{T-1} Q_t \cdot K_{t-\tau} \tag{9}$$

$Q$ and $K$ are the query and key vectors, and the delay $\tau$ represents the matching strength at different time offsets. This mechanism enhances the time series's ability to identify periodic and delayed relationships. The decomposed sum is simultaneously fed into the attention sub-network through its respective encoders. At the end of the encoder, this paper applies a fusion module to integrate the modeling results of the two sub-paths. Considering the trend encoding output as $Z_T$ and the perturbation encoding output as $Z_S$:

$$Z_T = Enc_{trend}(T_t) \tag{10}$$

$$Z_S = Enc_{seasonal}(S_t) \tag{11}$$

Then, performing dynamic weight fusion:

$$Z = \alpha \cdot Z_T + (1-\alpha) \cdot Z_S \tag{12}$$

$\alpha$ is a learnable parameter that adaptively allocates trends based on the time series variance. The final structure is formed and input into the decoder, which outputs the final predicted demand value through inverse reconstruction and dynamic frequency modulation.

### *C. Multi-scale Attention Mechanism for Supply Chain Management*

To capture multi-level cyclical patterns in logistics demand, such as daily, monthly, and quarterly patterns, this paper adds a nested multi-scale attention module after the Autoformer encoder, as shown in Fig. 2.

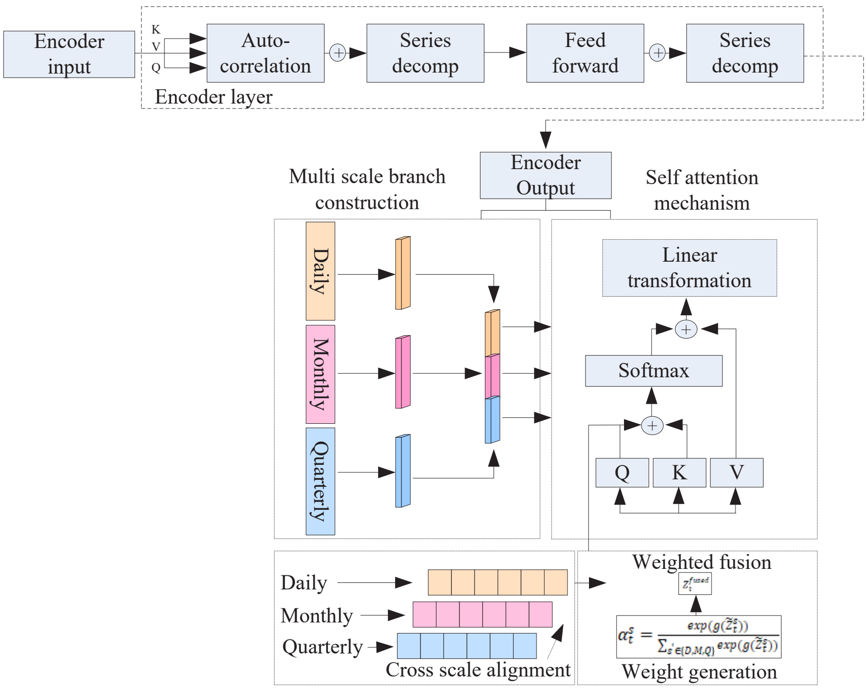


Fig. 2. Multi-scale Attention Module

This paper first reconstructs the long-term trend term $T_t$ and high-frequency residual term $S_t$ output by the decomposition module through multi-scale windows to construct input blocks of varying granularity. Assuming the original sequence length is $L$, three scale windows are defined:

1) Quarterly scale $s_Q$, window length $l_Q$;
2) Monthly scale $s_M$, window length $l_N$;
3) Daily scale $s_D$, window length $l_D$.

Let the intermediate representation after the trend decomposition module and encoder processing be $H \in R^{T \times N}$.

Based on the sliding window mechanism, the attention input sequences of three sub-paths are constructed respectively at the day level ($D$), month level ($M$), and quarter level ($Q$):

$$H^D=Segment(H,l_D)=\{H_1^D,H_2^D,H_3^D\} \tag{13}$$

$$H^M=Segment(H,l_M)=\{H_1^M,H_2^M,H_3^M\} \tag{14}$$

$$H^Q=Segment(H,l_Q)=\{H_1^Q,H_2^Q,H_3^Q\} \tag{15}$$

Then, corresponding attention modules are constructed for the inputs at the three scales. Each module adopts an improved local-global nested attention structure to capture both local perturbations and long-term dependencies. On each subpath, an independent multi-head self-attention module is used to extract local periodic dependency features:

$$Z^D=MSA_D(H^D) \tag{16}$$

$$Z^M=MSA_M(H^M) \tag{17}$$

$$Z^Q=MSA_Q(H^Q) \tag{18}$$

*MSA* represents a multi-head attention mechanism based on temporal position embedding to capture local dynamic patterns. Based on the periodic pattern extraction strategy at each scale $s_k$, a global representation across periods is constructed:

$$P_{s_k}=\sum_{j=0}^{\lfloor L/l_k\rfloor-1} X_{s_k}^{(jl_k)}\in\mathrm{R}^{\lfloor L/l_k\rfloor\times l_k\times N} \tag{19}$$

Then, building a cycle attention mechanism on the cycle dimension:

$$Attention_{global}^{(k)}=Softmax\left(\frac{Q_pK_p^{\mathrm{T}}}{\sqrt{d_k}}\right)V_p \tag{20}$$

In Formula (20), $Q_p$, $K_p$, and $V_p$ are the query, key, and value expressions in the periodic block. Due to the different sliding window lengths and step sizes of different scale paths, the lengths of the output sequences are different [25], [26]. To achieve subsequent fusion, the output features of different scales are length-aligned and reconstructed. Assuming the final target sequence length is $T'$, linear interpolation is used to map $Z^D$, $Z^M$, and $Z^Q$ back to the unified time axis:

$$\tilde{Z}^s=Align(Z^s,T'),\ s\in\{D,M,Q\} \tag{21}$$

*Align* represents the scale alignment operation, and outputs the aligned multi-scale feature tensor $\tilde{Z}^s\in\mathrm{R}^{T'\times N}$. A learnable scale fusion attention weight mechanism is employed to model the significance of features at different time scales. For each moment $t\in[1,T']$, three scale fusion weights are generated through the gating network $g(\cdot)$ [27], [28], [29]:

$$\alpha_t^s=\frac{exp(g(\tilde{Z}_t^s))}{\sum_{s'\in\{D,M,Q\}}exp(g(\tilde{Z}_t^{s'}))},\ s\in\{D,M,Q\} \tag{22}$$

The aligned multi-scale feature representations are weighted and fused to generate the final integrated sequence output $Z^{fused}\in\mathrm{R}^{T'\times N}$ [30], [31]:

$$Z_t^{fused}=\alpha_t^D\cdot\tilde{Z}_t^D+\alpha_t^M\cdot\tilde{Z}_t^M+\alpha_t^Q\cdot\tilde{Z}_t^Q,\ t\in[1,T'] \tag{23}$$

The output is fed into the decoder part of the Autoformer for final prediction reconstruction and abnormal residual scoring.

*D. Dynamic Early Warning Driven by Supply Chain Risks*

Based on model prediction errors, this method utilizes dynamic residual distribution modeling and multi-dimensional threshold adjustment strategies to score, identify, and trigger warning signals for abnormal conditions in system operation, providing data-driven risk prevention capabilities for logistics management. After the Autoformer model completes the modeling of trend and disturbance terms, these two types of features are collaboratively restored in the decoding phase to obtain the final prediction value sequence. Assuming that the trend prediction term output by the model is $\hat{T}_t$ and the disturbance prediction term is $\hat{S}_t$, the final demand forecast sequence can be written as:

$$\hat{y}_t=\hat{T}_t+\hat{S}_t,\ t=1,2,\cdots,\hat{H} \tag{24}$$

By retaining the independence of the characteristics of different time dynamic components and aggregating long-term trends and short-term disturbances, the demand forecast expression is achieved. Based on the historical true value $y_t$ and the predicted value $\hat{y}_t$, the time series residual vector $e_t$ is calculated:

$$e_t=y_t-\hat{y}_t \tag{25}$$

This residual sequence is directly used in the subsequent AD scoring mechanism, and its statistical characteristics are used to characterize the degree of deviation between normal and abnormal states. In response to the non-stationary and dynamic structural changes of SC data, this paper adopts a residual-driven dynamic anomaly scoring method. Let $e_t$ be the prediction residual at time point $t$, and $W$ be the size of the sliding window. The scoring function is defined based on the absolute deviation of the standardized residual:

$$z_t=\frac{e_t-\mu_e}{\sigma_e} \tag{26}$$

$\mu_e$ and $\sigma_e$ are the mean and standard deviation of the residual sequence $\{e_{t-W+1},\cdots,e_t\}$ in the sliding window, respectively:

$$\mu_e=\frac{1}{W}\sum_{i=t-W+1}^{t}e_i \tag{27}$$

$$\sigma_e=\sqrt{\frac{1}{W}\sum_{i=t-W+1}^{t}\left(e_t-\mu_e\right)^2} \tag{28}$$

The residual score function is defined as:

$$S_t=|z_t| \tag{29}$$

Setting the signal trigger threshold $\theta_t$ to:

$$\theta_t=\mu_{S_t}+K\cdot\sigma_{S_t} \tag{30}$$

If the current score meets the following conditions, it is considered abnormal:

$$S_t>\theta_t\Longrightarrow Anomaly\ at\ t \tag{31}$$

To improve the ability to identify multiple types of anomalies, this paper further designs multi-level judgment rules, combined time continuity and anomaly level distribution, and constructs an AW system based on a scoring function and duration, as shown in Table II:

TABLE II
ABNORMAL WARNING SYSTEM

| Abnormal level | Residual rating threshold | Duration | Response Level |
|---|---|---|---|
| Normal | $S_t \leq \theta_t$ | No requirements | No alarm |
| 1-Mild abnormality | $\theta_t < S_t \leq \theta_t + \delta$ | Continuous ≥ 2 time points | Yellow light reminder |
| 2- Moderate abnormality | $\theta_t + \delta < S_t \leq \theta_t + 2\delta$ | Continuous ≥ 2 time points | Orange light warning |
| 3- Severe abnormality | $S_t > \theta_t + 2\delta$ | Continuous ≥ 2 time points | Red light alarm |

In Table II, $\delta$ is the threshold increment step, which is used to construct the anomaly level interval in a hierarchical manner; the duration refers to the shortest continuous time required for the anomaly score to meet the conditions, which is used to enhance the stability of AD and avoid false alarms.

## IV. EXPERIMENTAL VERIFICATION IN SUPPLY CHAIN SCENARIOS

### A. Experimental Setup

To verify the effectiveness of the Autoformer-based SCLs demand modeling and AW method proposed in this paper, this paper selects real operational data from a large port logistics company between 2020 and 2024 to construct an experimental dataset. Its structure is displayed in Table III.

TABLE III
DATASET STRUCTURE

| Data type | Feature Name | Time granularity | Total amount of data | Data sources |
|---|---|---|---|---|
| Historical order data | Daily order quantity | Day | 1827 | Order management system |
| Inventory data | Current inventory level | Day | 1827 | Resource planning system |
| Shipping time | Average shipping time | Hour | 43848 | Port dispatch platform |
| Market environment | CPI | Month/Day | 48 | National Bureau of Statistics, State Council holiday release |
| | PMI | | 48 | |
| | Holiday signs | | 1827 | |

To ensure the timeliness of the evaluation and avoid data leakage, this paper divides the dataset into three independent time periods in chronological order: 1) Training set: January 1, 2020 - December 31, 2022 (1096 days in total); 2) Validation set: January 1, 2023 - December 31, 2023 (365 days in total); 3) Test set: January 1, 2024 - December 31, 2024 (366 days in total). For handling missing values, linear interpolation is used for continuous variables, and forward imputation is used for discrete variables. In time alignment, the order date is used as the primary index, and multi-source data is aligned through outer joins to ensure one record per day. Z-score normalization is applied to continuous features, with the mean and standard deviation calculated based on the training set. In sequence construction, a sliding window method is used to construct input-output pairs, with a window length 30 and a prediction step size 7. Finally, weekday, month, and holiday markers are added, and sine-cosine positional encoding is performed. To establish a reliable benchmark for anomaly detection and evaluation, a rule-based labeling process is adopted, defining supply chain logistics anomalies as "significant deviations from normal operating patterns in demand or operations caused by external disturbances or internal failures." Anomaly identification employs a multi-level set of rules for screening: demand surges (marked as candidate anomalies if the average daily order volume fluctuates by more than ±30% of the past 30-day rolling average for two consecutive days), transportation delays (marked as candidate anomalies if the average transportation time exceeds two standard deviations of the historical average for the same period and lasts for at least 24 hours), and inventory backlogs (marked as candidate anomalies if the inventory turnover rate is below the 10th percentile of the historical average for three consecutive days). The results were then independently reviewed and confirmed by three experts in the field of port logistics. The Fleiss' Kappa coefficient of the review results was 0.76, indicating that the labeling has good reliability. The model parameter settings for this paper are demonstrated in Table IV.

TABLE IV
MODEL PARAMETER SETTINGS

| Module | Parameter | Specifications |
|---|---|---|
| Input/output | Sliding window length | 30 |
| | Prediction step size | 7 |
| | Number of fused features | 8 |
| | Dimension of target prediction value | 1 |
| Autoformer basic parameters | Encoder layers | 3 |
| | Decoder layers | 2 |
| | Attention head count | 8 |
| | Hidden layer dimension | 512 |
| Adaptive decomposition parameters | Trend extraction window | 25 |
| Trend extraction window | Multi-scale sliding window | [90, 30, 7] |
| Training parameters | Learning rate | 1.00E-04 |
| | Optimizer | Adam |
| | Dropout ratio | 0.1 |
| | Batch size | 64 |
| | Max epochs | 100 |

The model was implemented using PyTorch 1.13.1, with Adam as the optimizer and a learning rate of $1\times10^{-4}$. During training, a model checkpoint (including weights, optimizer state, and the current epoch) was saved every 10 epochs. The final model was selected based on the checkpoint with the lowest WAPE on the validation set. Training logs (loss and evaluation metrics) were saved as a comma separated values (CSV) file for subsequent analysis. To ensure reproducibility, all randomization processes used a fixed seed of 2024.

To comprehensively evaluate the effectiveness of the method, it selects three representative time series forecasting and AD models as baselines for comparison:

1) ARIMA: A classic linear model used for analyzing stationary time series.

2) LSTM: A recurrent neural network with long-term memory, used for modeling nonlinear time series;

3) Informer: A Transformer variant based on sparse attention, suitable for long time series forecasting.

Taking ARIMA, LSTM and Informer as control models, the effectiveness of this method in SCLs demand modeling and AW is evaluated from the two levels of LD prediction accuracy and AW effectiveness.

### *B. Supply Chain Demand Forecasting Performance*

SC demand forecasting emphasizes balancing a model's capacity to predict long-term trends with its responsiveness to local disturbances. To further analyze the model's performance in terms of forecast accuracy, this paper compares the WAPE and RMSE performance of various models at different forecast step sizes:

$$WAPE=\frac{\sum_{t=1}^{n}|y_t-\hat{y}_t|}{\sum_{t=1}^{n}|y_t|} \tag{32}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2} \tag{33}$$

Fig. 3 illustrates the outcomes.

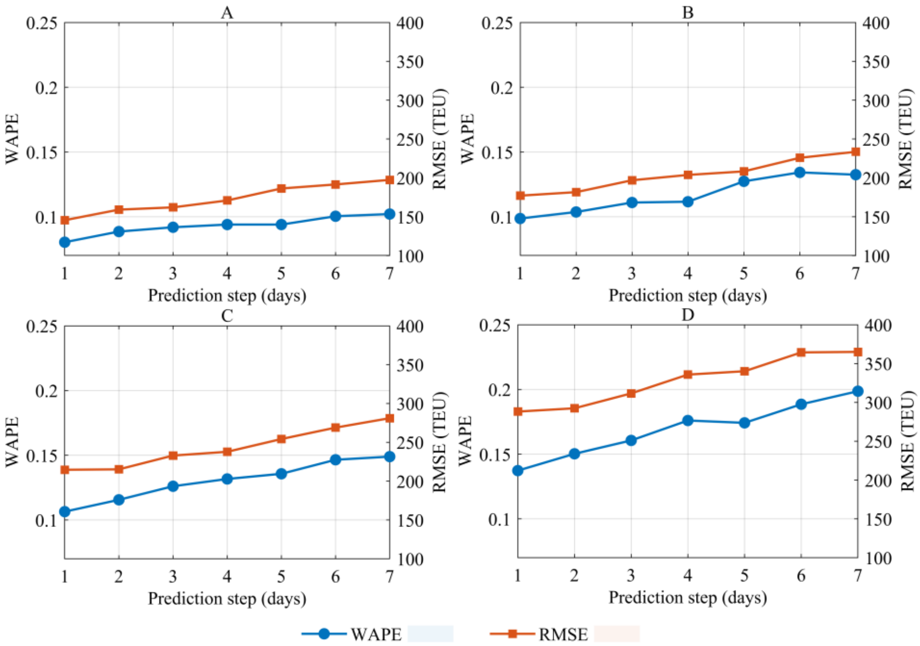


Fig. 3. WAPE and RMSE Comparison Results; A) shows the WAPE and RMSE of the Autoformer, B) shows the WAPE and RMSE of the Informer, C) shows the WAPE and RMSE of the LSTM, D) shows the WAPE and RMSE of the ARIMA.

The outcomes in Fig. 3 demonstrate that the proposed Autoformer-based SCLs demand modeling method outperforms the comparison methods in both WAPE and RMSE metrics. The trends in WAPE and RMSE over the forecast step size indicate that the forecast errors of all methods increase with increasing forecast step size, while Autoformer maintains the lowest forecast error over the 7-day forecast period. In Fig. 3A, its WAPE and RMSE on the seventh day are 0.105 and 200 TEU, respectively. The Autoformer's adaptive trend decomposition mechanism effectively separates long-term trends from short-term fluctuations in LD. Extracting cyclical features in parallel at multiple time scales, such as daily, monthly, and quarterly, fully exploits the multi-level temporal dependencies of PLD, resulting in stronger long-term forecast accuracy. However, the fixed time scale modeling method adopted by the Informer in Fig. 3B is difficult to adapt to the joint impact of multi-level periodic characteristics in PLD; although the LSTM in Fig. 3C has memory capabilities, it is prone to falling into local optimality and cannot effectively handle multi-scale time dependencies; The ARIMA model in Fig. 3D relies on linear assumptions, making it challenging to capture the nonlinear complex relationships and dynamic dependencies among multiple variables in PLD.

To further verify the performance advantage of this method in capturing the changing trend of SCLs demand, it applies a trend matching index. It uses the Spearman rank correlation coefficient within a sliding window to compute the consistency between the predicted and actual sequence in the direction of change. Then it compares and analyzes the performance of different methods in the trend matching index, as shown in Fig. 4.

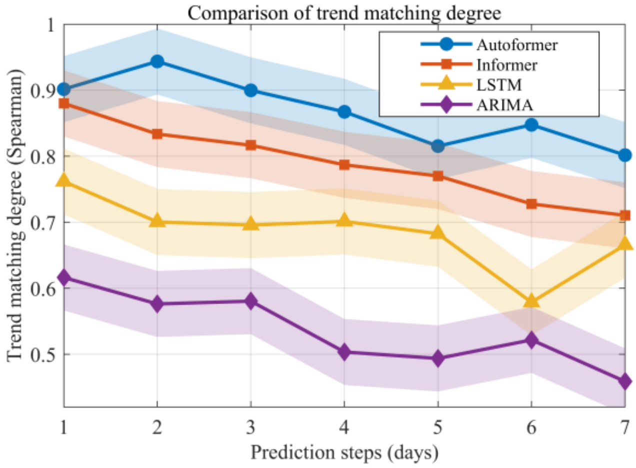


Fig. 4. Trend Matching Comparison Results

As shown in Fig. 4, the proposed Autoformer method generally achieves higher trend matching across different forecast steps. The correlation coefficient shows that the Autoformer achieves a peak trend matching score exceeding 0.900 within a seven-day forecast period, reaching 0.808 on the seventh day. Even with longer forecast steps, the Autoformer maintains high trend consistency. In contrast, the trend matching scores of the Informer, LSTM, and ARIMA methods decrease from initial values of 0.883, 0.768, and 0.615 to 0.712, 0.663, and 0.450, respectively, demonstrating their limitations in long-term trend forecasting. Based on multi-scale modeling and adaptive decomposition, this method decomposes the LD series into long-term trend terms and short-term disturbance terms, accurately modeling each frequency component separately. This effectively avoids trend ambiguity when processing complex time series. Compared to other methods, the Autoformer demonstrates superior trend matching in the complex and dynamic environment of PLD.

### C. *Verification of the Effectiveness of Supply Chain Anomaly Warning*

To assess how well the AW mechanism performs in SCLs scenarios, dynamic thresholds and multi-level judgment rules are set to statistically analyze the classification performance of each model at different anomaly levels. Three core indicators, precision, recall, and F1-score, were used for the quantitative evaluation. Table V displays the results.

TABLE V
COMPARISON OF PRECISION, RECALL, AND F1-SCORE

| Model | Abnormal level | Precision | Recall | F1-score |
|---|---|---|---|---|
| Autoformer | Mild abnormality | 0.882 | 0.857 | 0.869 |
| | Moderate abnormality | 0.835 | 0.901 | 0.867 |
| | Severe abnormality | 0.916 | 0.934 | 0.925 |
| Informer | Mild abnormality | 0.765 | 0.723 | 0.743 |
| | Moderate abnormality | 0.708 | 0.816 | 0.758 |
| | Severe abnormality | 0.852 | 0.829 | 0.840 |
| LSTM | Mild abnormality | 0.684 | 0.652 | 0.668 |
| | Moderate abnormality | 0.621 | 0.703 | 0.659 |
| | Severe abnormality | 0.789 | 0.751 | 0.770 |
| ARIMA | Mild abnormality | 0.543 | 0.587 | 0.564 |
| | Moderate abnormality | 0.502 | 0.624 | 0.556 |
| | Severe abnormality | 0.658 | 0.603 | 0.629 |

As shown in Table V, the proposed method obtains F1-scores of 0.869, 0.867, and 0.925, respectively. The proposed method performs particularly well in severe AD, with F1-score improvements of 10.1%, 20.1%, and 47.1% over the Informer, LSTM, and ARIMA methods, respectively. The traditional ARIMA method exhibits the lowest detection performance at all anomaly levels, primarily due to its linear modeling assumptions, which are hard to adapt to the complex nonlinear characteristics of port logistics data. While the LSTM method accounts for temporal dependencies, it has limitations when handling multivariate coordinated anomalies. While the Informer improves modeling capabilities through its attention mechanism, its ability to identify anomalies of varying severity is relatively limited.

To directly reflect the model's effectiveness in detecting and warning real risks, it uses FN to quantify the underreporting of each model at different anomaly levels. It selects supplier anomalies, inventory anomalies, logistics node anomalies, and market fluctuation anomalies from port logistics in 2023 as the test set. Fig. 5 demonstrated the outcomes.

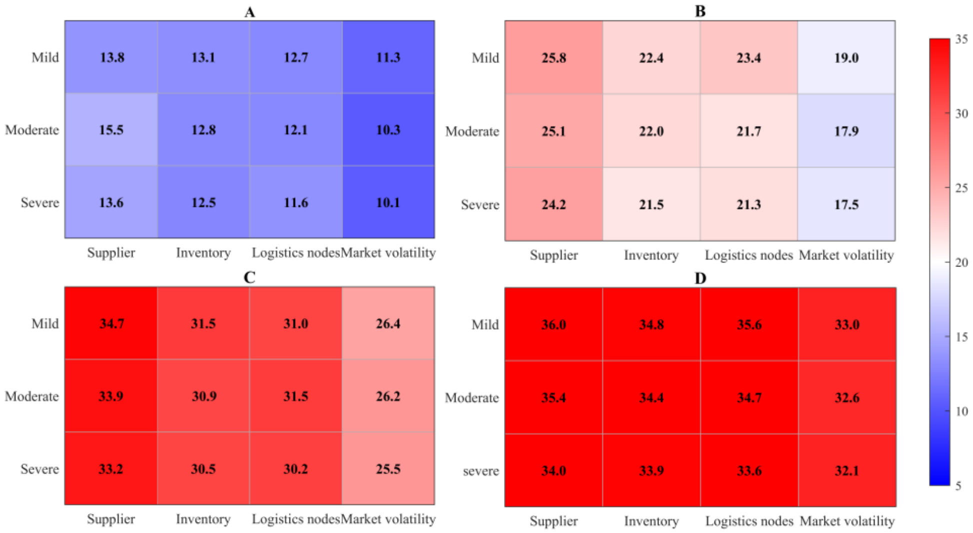


Fig. 5. FNR Comparison Results; A) shows the FNR of the Autoformer, B) shows the FNR of the Informer, C) shows the FNR of the LSTM, D) shows the FNR of the ARIMA

The results in Fig. 5 demonstrate that the Autoformer demonstrates significant advantages in detecting all four anomaly types. In the four severe anomaly scenarios, the average FNR of the Autoformer in Fig 5A is 12.0%, which is 9.1% lower than that of the Informer in Fig 5B (21.1%), 17.9% lower than that of the LSTM in Fig 5C (29.9%), and 21.4% lower than that of the ARIMA in Fig 5D (33.4%). This demonstrates that the Autoformer's adaptive trend decomposition effectively captures complex anomalies. The MSAM allows the model to detect anomaly features across various time scales, improving its capacity to recognize complex anomaly patterns. A dynamic thresholding mechanism based on prediction residuals avoids the limitations of traditional fixed threshold methods and enables adaptive detection of different types of anomalies. This synergistic effect makes the method more reliable in dealing with the complex and ever-changing anomaly scenarios in port logistics systems.

Further, the FPR is calculated to compare the proportion of standard data samples that are incorrectly identified as anomalies by different AD models to measure the specificity of the model in logistics early warning. Fig. 6 illustrates the findings.

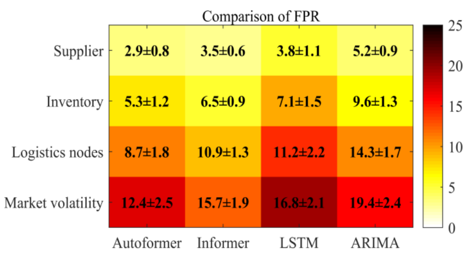


Fig. 6. FPR Comparison Results

As displayed in Fig. 6, for the four warning scenarios, the model in this paper achieves an average FPR of 7.3%; the Informer, LSTM, and ARIMA models achieve 9.2%, 9.7%, and 12.1%, respectively. Comparative results show that the Autoformer model outperforms the Informer, LSTM, and ARIMA models. In the four scenarios of SC, inventory, logistics nodes, and market fluctuations, short-term sharp fluctuations and cyclical structural misalignments can easily lead conventional models to misclassify them as anomalies. However, the model, through scale fusion and residual modeling, proactively accounts for these variations, significantly reducing the false alarm rate. Applying trend-disturbance decoupling at the encoding stage prevents short-term fluctuations from interfering with long-term trend modeling. The MSAM further improves the model's accuracy in discerning anomaly boundaries at different time series frequencies, resulting in more favorable FPR results and effectively improving warning performance.

Further, the average detection delay of each model was compared, and the time difference from the appearance of abnormal features in the input data to the output of the warning signal by the model was recorded. The findings of each model's statistical analysis, based on the backtest of abnormal event samples, are presented in Table VI.

TABLE VI
AVERAGE DETECTION LATENCY COMPARISON

| Model | Mild abnormality (h) | Moderate abnormality (h) | Severe abnormality (h) | Overall average (h) |
|---|---|---|---|---|
| Autoformer | 1.8±0.3 | 1.2±0.2 | 0.9±0.2 | 1.3±0.4 |
| Informer | 3.2±0.6 | 2.7±0.5 | 2.1±0.4 | 2.7±0.8 |
| LSTM | 4.1±0.9 | 3.5±0.7 | 2.8±0.6 | 3.5±1.1 |
| ARIMA | 6.5±1.5 | 5.3±1.2 | 4.2±1.0 | 5.3±1.6 |

Table VI shows that the four models exhibit varying detection latency: Autoformer has the lowest overall average latency, reaching 1.3 ± 0.4 hours, while ARIMA has the highest, reaching 5.3 ± 1.6 hours. Informer and LSTM have average overall latencies of 2.7 ± 0.8 hours and 3.5 ± 1.1 hours, respectively. The comparative results demonstrate that Autoformer's adaptive decomposition and residual-driven early warning mechanism enable it to maintain a relatively fast response in AD.

To further validate the model's practicality and reliability in logistics anomaly early warning tasks, a weighted Kappa coefficient is used to measure the consistency between predictions and true levels. Key performance indicators, such as the over-threshold deviation rate and severe anomaly recognition rate, are also statistically analyzed, as displayed in Fig. 7.

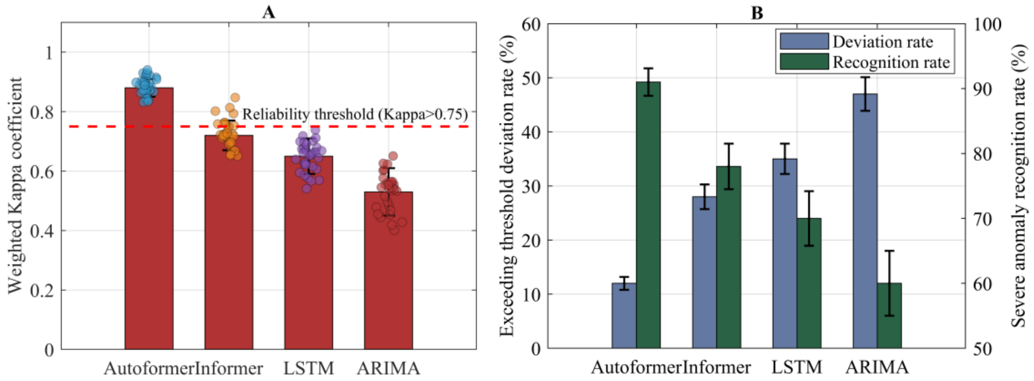


Fig. 7. Comparison of Abnormal Warning Level Consistency; A) shows the weighted Kappa coefficient, B) shows the key performance indicators

As shown in Fig. 7, the Autoformer demonstrates superior performance across all metrics. The weighted Kappa coefficient in Fig. 7A is as high as 0.844, significantly exceeding the reliability threshold of 0.75, indicating a high degree of consistency between its predicted grades and the ground truth. The Informer performs second (Kappa = 0.728), while the Kappa coefficients for the LSTM and ARIMA models are 0.650 and 0.537, respectively. Among the key performance indicators in Fig. 7B, the Autoformer achieves the lowest over-threshold deviation rate (12%) and a severe anomaly recognition rate exceeding 90%, demonstrating not only overall accurate predictions but also effective warning of critical, high-risk events in SCLs. In contrast, the ARIMA model exhibits the weakest performance, with a deviation rate of 47% and a recognition rate of only 60%. While the LSTM model demonstrates some recognition capability, it suffers from significant deviation. The Informer performs moderately well in complex models, outperforming the LSTM and ARIMA models but not the Autoformer. Overall, the Autoformer demonstrates superior performance in terms of grade prediction accuracy, stability, and key event capture, making it suitable for AW in highly dynamic and volatile logistics environments.

### D. *Ablation Experiment*

To verify the contribution of each module to the model performance, an ablation experiment was designed. Related components were removed or replaced sequentially, and their performance was evaluated under the same experimental settings. Six model variants were constructed for comparison: 1) Complete model (Proposed); 2) Quarterly branch removed: only daily and monthly scale attention branches were retained; 3) Monthly branch removed: only daily and quarterly scale attention branches were retained; 4) Dual-path modeling removed: the original Autoformer fixed decomposition method was used instead of the adaptive trend-perturbation dual-path structure; 5) Static Z-score thresholding: the dynamic residual scoring and multi-level thresholding were replaced with the traditional static Z-score method; 6) Single path + static thresholding only: both monthly and quarterly branches in the multi-scale attention were removed, and a static thresholding method was used. The performance comparison of each variant on the two tasks of demand forecasting and anomaly warning is shown in Table VII.

TABLE VII
ABLATION TEST RESULTS

| MODEL | WAPE | RMSE (TEU) | F1 (SERIOUS ABNORMALITY) | FNR (%) | FPR (%) | KAPPA |
|---|---|---|---|---|---|---|
| FULL MODEL (PROPOSED) | 0.105 | 200 | 0.925 | 12 | 7.3 | 0.844 |
| REMOVE QUARTERLY BRANCHES | 0.118 | 225 | 0.891 | 15.2 | 8.1 | 0.812 |
| REMOVE MONTHLY BRANCHES | 0.121 | 231 | 0.885 | 16 | 8.5 | 0.806 |
| FIXED DECOMPOSITION | 0.129 | 245 | 0.872 | 18.7 | 9 | 0.781 |
| STATIC THRESHOLD | 0.107 | 203 | 0.868 | 17.5 | 10.4 | 0.765 |
| SINGLE PATH + STATIC THRESHOLD | 0.135 | 252 | 0.821 | 22.3 | 11.8 | 0.732 |

As shown in Table VII, removing both quarterly and monthly branches led to an increase in WAPE and RMSE, and a decrease in F1, indicating that quarterly and monthly scales are crucial for capturing medium- to long-term cyclical fluctuations. Fixed decomposition significantly outperformed the full model across all metrics, demonstrating that adaptive decomposition more effectively decouples trends from disturbances and improves the model's ability to model non-stationary sequences. While static thresholds were close to the full model in prediction error, they performed significantly worse in anomaly detection, indicating that dynamic thresholds and multi-level judgments significantly improve the ability to distinguish between anomalies of different degrees and reduce false alarms. The simplified structure and static thresholds exhibited the worst performance, suggesting a synergistic enhancement effect among the modules.

### *E. Comparison of SOTA Methods*

To comprehensively evaluate the effectiveness of this method, five representative time series forecasting and anomaly detection models were selected for comparison to ensure the comparison's sophistication and fairness. Each model was implemented using its official codebase and employed the same data partitioning, preprocessing procedures, and evaluation protocol as this method.

1) PatchTST is a Transformer architecture based on channel-independent patches, exhibiting excellent performance in long-term prediction tasks.

2) TimesNet transforms time series data into a two-dimensional space to model multi-period features, widely regarded as a general-purpose time series model.

3) iTransformer employs an inverted attention mechanism, achieving significant performance in multivariate time series prediction.

4) Autoformer (original) is a classic time series Transformer model using autocorrelation decomposition, serving as the foundational version of this method.

The anomaly detection performance comparison is shown in Table VIII:

TABLE VIII
PERFORMANCE COMPARISON OF SOTA METHODS FOR ANOMALY DETECTION

| MODEL | F1 (SERIOUS ABNORMALITY) | F1 (MINOR ABNORMALITY) | FNR (%) | FPR (%) | KAPPA |
|---|---|---|---|---|---|
| PROPOSED | 0.925 | 0.869 | 12.0 | 7.3 | 0.844 |
| PATCHTST | 0.891 | 0.832 | 15.6 | 9.1 | 0.815 |
| TIMESNET | 0.902 | 0.841 | 14.8 | 8.7 | 0.826 |
| ITRANSFORMER | 0.898 | 0.838 | 15.2 | 9.0 | 0.819 |
| ORIGINAL AUTOFORMER | 0.867 | 0.812 | 18.7 | 9.5 | 0.781 |

As shown in Table VIII, compared with PatchTST, TimesNet, iTransformer, and the original Autoformer, the model in this paper has a greater advantage in anomaly detection performance, with a severe anomaly F1 score of over 0.9, specifically 0.925. Except for TimesNet (0.902), the other models all have scores below 0.9, and their prediction errors are all higher than that of the model in this paper. Overall, the proposed method performs better in terms of prediction accuracy, anomaly detection capability, and early warning consistency, verifying the effectiveness and advancement of multi-scale attention, adaptive decomposition, and dynamic threshold mechanisms in complex time-series scenarios of port logistics.

### *F. Parameter Sensitivity Analysis*

To verify the robustness of the critical design choices, we conducted sensitivity analyses on the sliding window length, multi-scale window length, and threshold coefficient. These are shown in Tables IX, 10, and 11.

TABLE IX
SLIDING WINDOW LENGTH SENSITIVITY ANALYSIS

| DAYS | WAPE | RMSE | F1 (SERIOUS ABNORMALITY) | TRAINING TIME (EPOCH/MIN) |
|---|---|---|---|---|
| 15 | 0.121 | 228 | 0.901 | 3.2 |
| 30 | 0.105 | 200 | 0.925 | 4.5 |
| 45 | 0.108 | 205 | 0.918 | 6.1 |
| 60 | 0.113 | 212 | 0.91 | 8.3 |

As shown in Table IX, when the sliding window length is 30, the model achieves near-optimal performance in both WAPE and F1 scores. Too short a window (15 days) leads to a decrease in predictive performance due to insufficient historical information, while too long a window (60 days) introduces excessive noise and significantly increases computational costs.

TABLE X
MULTISCALE WINDOW LENGTH SENSITIVITY ANALYSIS

| COMBINATION | TREND MATCHING DEGREE | F1 (SERIOUS ABNORMALITY) | KAPPA |
|---|---|---|---|
| (20, 60) | 0.861 | 0.903 | 0.811 |
| (30, 90) | 0.9 | 0.925 | 0.844 |
| (40, 120) | 0.878 | 0.915 | 0.829 |

When the monthly window is 30 and the quarterly window is 90, which aligns with the natural business cycle, the model achieves optimal performance in trend matching, severe anomaly detection (F1), and Kappa. This indicates that defining multi-scale windows based on natural cycles can more effectively extract time-series patterns that correspond to the decision-making level of logistics management. Windows that are too long or too short will lead to feature extraction bias and performance degradation.

TABLE XI
SENSITIVITY ANALYSIS OF DYNAMIC THRESHOLD COEFFICIENT

| THRESHOLD | F1 (SERIOUS ABNORMALITY) | FNR (%) | FPR (%) |
|---|---|---|---|
| 1.5 | 0.938 | 9.2 | 12.6 |
| 2 | 0.925 | 12 | 7.3 |
| 2.5 | 0.901 | 16.5 | 5.1 |
| 3 | 0.872 | 21.3 | 3.8 |

When the threshold is 2.0, the model achieves the best balance between FPR and FNR, and has the highest overall score. Lowering the threshold to 1.5, while capturing more anomalies, significantly increases false alarms and reduces warning specificity; raising the threshold above 2.5 leads to a substantial increase in missed alarms. This verifies that choosing a threshold of 2.0 can maintain overall stability while ensuring warning sensitivity.

### G. *Method Generalization Ability Analysis*

To explore the applicability of the proposed method in relevant scenarios, supplementary experiments were conducted on a publicly available supply chain dataset. The model structure remained unchanged; only the input dimensions were adjusted based on the data features, and the model was retrained. The results are shown in Table XII.

TABLE XII
GENERALIZATION ABILITY ANALYSIS

| MODEL | WAPE | RMSE | F1 (SERIOUS ABNORMALITY) | TRAINING TIME (EPOCH/MIN) |
|---|---|---|---|---|
| PATCHTST | 0.183 | 415 | 0.812 | 5.1 |
| TIMESNET | 0.176 | 398 | 0.826 | 6.3 |
| PROPOSED | 0.162 | 372 | 0.845 | 7.2 |

Table XII shows that the proposed method maintains relatively ideal overall performance without requiring specific structural adjustments to the dataset. This initially indicates that the proposed multi-scale fusion, adaptive decomposition, and dynamic early warning mechanism has a certain degree of generality for logistics time-series data with similar periodicity and non-stationarity. However, its effectiveness in scenarios with significant differences still needs further verification. Future work should further develop a lighter and more adaptive architecture to improve cross-domain generalization capabilities.

## V. CONCLUSION

Taking into full account the support capabilities of modern information and communication infrastructure for real-time intelligent decision-making, this paper proposes a deep time series modeling framework for cyber-physical fusion supply chain systems. This paper addresses the unique complexity of port logistics time-series data by systematically integrating and improving self-transformation decomposition, multi-scale attention, and dynamic residual scoring mechanisms through scenario-oriented architecture integration and optimization, in order to decouple trends and disturbances in cargo throughput. Furthermore, it constructs an anomaly scoring function using a residual-driven dynamic threshold mechanism to improve the sensitivity and robustness of the warning system. Experimental results show that, within a 7-day forecast period, the proposed model achieves relatively lower WAPE and RMSE in cargo throughput forecasting compared to the baseline. Its performance on trend matching indicators demonstrates its good trend-following ability. Through multi-scale modeling, trend construction, and anomaly identification, this paper provides a certain reference for intelligent risk perception and response in high-volatility port logistics scenarios.

## COMPETING INTERESTS

No competing interests are disclosed by the authors.

## AUTHORSHIP CONTRIBUTION STATEMENT

Xizhuo Han: Supervision, Conceptualization, Project administration, Writing-Original draft preparation.

Chuanchuan Teng: Validation, Methodology, Software.

## DATA AVAILABILITY

Available upon request.

## DECLARATIONS

Not applicable.

## CONFLICTS OF INTEREST

The writers assert that they have no conflict of interest pertaining to the release of this article.

## AUTHOR STATEMENT

All authors have reviewed and endorsed the manuscript, confirming it fulfills the previously mentioned criteria for authorship. Furthermore, each author is confident that the manuscript accurately reflects genuine research.

## FUNDING

Not applicable.

## ETHICAL APPROVAL

Each author was directly and significantly involved in the work that resulted in this paper, and they will all accept public accountability for what it contains.

## REFERENCES

[1] A. Aamer, L. Eka Yani, and Im. Alan Priyatna, "Data analytics in the supply chain management: Review of machine learning applications in demand forecasting," *Operations and Supply Chain Management: An International Journal*, vol. 14, no. 1, pp. 1–13, 2020.

[2] K. Douaioui, R. Oucheikh, O. Benmoussa, and C. Mabrouki, "Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review.," *Applied System Innovation (ASI)*, vol. 7, no. 5, 2024. DOI: 10.3390/asi7050093

[3] M. R. Bhuiyan, S. I. Rohan, F. Rahman Sheikh, and M. S. Alam, "Supply Chain Risk Management: Strategic Solutions for Reducing Transportation and Logistics Risks," *Academic Journal of Innovation, Engineering & Emerging Technology*, vol. 1, no. 1, pp. 72–90, 2024. DOI: 10.69593/ajieet.v1i01.125

[4] H. Panjehfouladgaran and S. F. W. T. Lim, "Reverse logistics risk management: identification, clustering and risk mitigation strategies," *Management Decision*, vol. 58, no. 7, pp. 1449–1474, 2020. DOI: 10.1108/MD-01-2018-0010

[5] K. Lai, M. Vejvar, and V. Y. H. Lun, "Risk in port logistics: Risk classification and mitigation framework," *International Journal of Shipping and Transport Logistics*, vol. 12, no. 6, pp. 576–596, 2020. DOI: 10.1504/IJSTL.2020.111117

[6] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Adv Neural Inf Process Syst*, vol. 34, pp. 22 419–22 430, 2021.

[7] L. Li, Y. Gong, Z. Wang, and S. Liu, "Big data and big disaster: a mechanism of supply chain risk management in global logistics industry," *International Journal of Operations & Production Management*, vol. 43, no. 2, pp. 274–307, 2023. DOI: 10.1108/IJOPM-04-2022-0266

[8] S. Yerra, "Optimizing supply chain efficiency using AI-driven predictive analytics in logistics," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 11, no. 2, pp. 1212–1220, 2025. DOI: 10.32628/CSEIT25112475

[9] B. He and L. Yin, "Prediction modelling of cold chain logistics demand based on data mining algorithm," *Math Probl Eng*, vol. 2021, no. 1, p. 342 1478, 2021. DOI: 10.1155/2021/3421478

[10] L. Terrada, M. El Khaili, and H. Ouajji, "Demand forecasting model using deep learning methods for supply chain management 4.0," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022.

[11] J. Zhu, Y. Wu, Z. Liu, and C. Costa, "Sustainable optimization in supply chain management using machine learning," *International Journal of Management Science Research*, vol. 8, no. 1, pp. 1–8, 2025.

[12] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep isolation forest for anomaly detection," *IEEE Trans Knowl Data Eng*, vol. 35, no. 12, pp. 12 591–12 604, 2023. DOI: 10.1109/TKDE.2023.3270293

[13] Y. Chen, H. Zhang, X. Yan, and Q. Miao, "Supply chain demand forecasting based on multi-time scale data fusion network," *Comput Ind Eng*, p. 111 324, 2025. DOI: 10.1016/j.cie.2025.111324

[14] X. I. E. Luqiang, X. U. Jing, and W. Haiyan, "Risk early warning model of cruise ship construction material logistics collection and distribution based on RS-RBFNN," *China Safety Science Journal*, vol. 33, no. 6, p. 114, 2023.

[15] X. Ma and H. Zhang, "Time Series Forecasting Method Based on Multi- Scale Feature Fusion and Autoformer," *Applied Sciences*, vol. 15, no. 7, p. 3768, 2025. DOI: 10.3390/app15073768

[16] D. Cao and S. Zhang, "AD-autoformer: decomposition transformers with attention distilling for long sequence time-series forecasting," *J Supercomput*, vol. 80, no. 14, pp. 21 128–21 148, 2024. DOI: 10.1007/s11227-024-06266-8

[17] Z. Tao, Q. Xu, X. Liu, and J. Liu, "An integrated approach implementing sliding window and DTW distance for time series forecasting tasks," *Applied Intelligence*, vol. 53, no. 17, pp. 20 614–20 625, 2023. DOI: 10.1007/s10489-023-04590-9

[18] M. A. Jahin, A. Shahriar, and M. Al Amin, "MCDFN: supply chain demand forecasting via an explainable multi-channel data fusion network model," *Evol Intell*, vol. 18, no. 3, p. 66, 2025. DOI: 10.1007/s12065-025-01053-7

[19] W. Zheng *et al.*, "Design of a modified transformer architecture based on relative position coding," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 168, 2023. DOI: 10.1007/s44196-023-00345-z

[20] N. M. Foumani, C. W. Tan, G. I. Webb, and M. Salehi, "Improving position encoding of transformers for multivariate time series classification," *Data Min Knowl Discov*, vol. 38, no. 1, pp. 22–48, 2024. DOI: 10.48550/arXiv.2305.16642

[21] C. Duan, G. Xiu, and Y. Zhang, "Coordinated management method of information contract in port logistics service supply chain," *J Coast Res*, vol. 93, no. SI, pp. 1047–1052, 2019. DOI: 10.2112/SI93-151.1

[22] L. Shuo, H. Zehua, W. Shuo, Y. Qian, and L. Wenbo, "Research on Enterprise Logistics Management Based on Big Data," in *2024 Second International Conference on Data Science and Information System (ICDSIS)*, IEEE, 2024, pp. 1–4. DOI: 10.1109/ICDSIS61070.2024.10594210

[23] Y. Chen, W. Jia, and Q. Wu, "Fine-scale deep learning model for time series forecasting," *Applied Intelligence*, 2024. DOI: 10.1007/s10489-024-05701-w

[24] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep learning for time series forecasting: a survey," *Big Data*, vol. 9, no. 1, pp. 3–21, 2021. DOI: 10.1089/big.2020.0159

[25] Q. Yu, G. Yang, X. Wang, Y. Shi, Y. Feng, and A. Liu, "A review of time series forecasting and spatio-temporal series forecasting in deep learning," *J Supercomput*, vol. 81, no. 10, pp. 1–48, 2025. DOI: 10.1007/s11227-025-07632-w

[26] G. Tian, C. Zhang, Y. Shi, and X. Li, "MultiWaveNet: A long time series forecasting framework based on multi-scale analysis and multi-channel feature fusion," *Expert Syst Appl*, vol. 251, p. 124 088, 2024. DOI: 10.1016/j.eswa.2024.124088

[27] G. Su and Y. Guan, "MSDformer: an autocorrelation transformer with multiscale decomposition for long-term multivariate time series forecasting," *Applied Intelligence*, vol. 55, no. 3, p. 179, 2025. DOI: 10.1007/s10489-024-06105-6

[28] S. Feng *et al.*, "Multi-scale attention flow for probabilistic time series forecasting," *IEEE Trans Knowl Data Eng*, vol. 36, no. 5, pp. 2056–2068, 2023. DOI: 10.1109/TKDE.2023.3319672

[29] C. Lei, H. Zhang, Z. Wang, and Q. Miao, "Multi-Model Fusion Demand Forecasting Framework Based on Attention Mechanism," *Processes*, vol. 12, no. 11, p. 2612, 2024. DOI: 10.3390/pr12112612

[30] J. Li, L. Zhu, Y. Zhang, D. Guo, and X. Xia, "Attention-based multi-scale prediction network for time-series data," *China Communications*, vol. 19, no. 5, pp. 286–301, 2021. DOI: 10.23919/JCC.2021.00.005

[31] J. Lu, L. Pan, and Q. Ren, "Spatial-temporal memory enhanced multi-level attention network for origin-destination demand prediction," *Complex & Intelligent Systems*, vol. 10, no. 5, pp. 6435–6448, 2024. DOI: 10.1007/s40747-024-01494-0

**Chuanchuan Teng** graduated from Dalian Maritime University in 2016 with a Master's degree in logistics engineering and management, and now he is working in Cangzhou Jiaotong College. His research interest includes supply chain management and port logisitics.

**Xizhuo Han** graduated from Yanshan University in 2020 with a Master's degree in logistics engineering, and now she is working in Cangzhou Jiaotong College. Her research interest includes supply chain management and port logisitics.