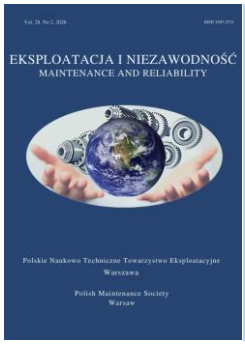


Eksploatacja i Niezawodnosc – Maintenance and Reliability
Volume 28 (2026), Issue 2
journal homepage: http://www.ein.org.pl

Article citation info:
Huang X, Wang Y, Yue S, Wang J, Han Y, Spare parts consumption prediction model for improving maintenance and operational reliability, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(2) http:/10.17531/ein/210721

Spare parts consumption prediction model for improving maintenance and operational reliability

Indexed by: Web of Science Group

Xueding Huang[a], Yabin Wang[a,*], Shuai Yue[b] Jinguo Wang[a], Yujian Han[a]

[a] Department of Sixth Research, Shijiazhuang Campus, Army Engineering University, China
[b] Shijiazhuang Campus, Army Engineering University, Shijiazhuang 050003, China

Highlights

- Spare Parts Consumption Prediction.
- Reliability Engineering.
- Maintenance Reliability.
- Hybrid Model.

Abstract

In fields such as industrial production, the reliability of mechanical equipment maintenance is affected by the inventory management of maintenance spare parts. Accurately predicting the consumption of maintenance spare parts is of great significance for optimizing resource allocation and formulating scientific maintenance strategies. However, spare parts consumption is often comprehensively affected by various nonlinear and multi-scale factors, and the existing prediction methods are difficult to effectively capture these complex characteristics. To address this issue, this paper proposes a hybrid prediction model integrating STL decomposition, iTransformer and TimesNet. This model combines time series decomposition technology with deep learning frameworks and is capable of efficiently handling long-term series data and their periodic characteristics. Based on the ten-year continuous historical consumption data of spare parts in a certain warehouse, the empirical results show that this hybrid model significantly outperforms multiple benchmark models in multiple key performance indicators. The precise prediction ability of this method is conducive to improving the scientific nature of spare parts management, thereby enhancing the reliability of the overall maintenance system, which is of great practical significance for enterprises to improve the level of equipment maintenance.

Keywords
spare parts consumption prediction，reliability engineering，maintenance reliability，hybrid model，STL decomposition，iTransformer，TimesNet

1. Introduction

In modern industrial production, predictive maintenance, as an effective maintenance activity to maintain the operational reliability and availability of production equipment and systems, is a frequently adopted equipment maintenance strategy by enterprises123. As a core element of maintenance activities, the level of spare parts management directly affects maintenance efficiency and system reliability4. Timely and accurate acquisition of required maintenance spare parts is the key to ensuring reliable maintenance activities and reducing downtime. Failure to effectively predict and manage spare parts inventory

(*) Corresponding auth
E-mail addresses: X. Huang (ORCID:0009-0005-7387-0431) 745869300@qq.com, Y. Wang(ORCID:0000-0002-7568-3518) wangyabin123@163.com, S. Yue (ORCID:0000-0003-4743-7202) 13313216227@163.com J. Wang (ORCID:0009-0006-7394-9008) jinguowang22@163.com, Y. Han (ORCID:0009-0003-4176-7320) 358721922@qq.com

may result in a shortage of spare parts, causing maintenance delays or even long-term equipment shutdowns, seriously affecting the operational reliability and availability of the system. And with the widespread application of AI driven new computing methods in the fields of materials science and equipment engineering 5, these studies not only drive technological innovation, but also indicate that future equipment and systems will be more complex and intelligent, and their spare parts may exhibit higher integration, longer lifespan, and more complex failure modes. Therefore, accurate and reliable prediction of maintenance spare parts consumption is of great significance for developing scientific maintenance plans and optimizing inventory management, thereby enhancing the reliability of maintenance spare parts guarantee. However, spare parts consumption data often exhibits significant complex characteristics such as nonlinearity, intermittency, multi-scale periodicity, and uncertainty, making it difficult to capture their consumption patterns. Traditional prediction methods often have limitations in handling these complex time series, making it difficult to provide high-precision and robust prediction results that meet the requirements of reliable maintenance decisions. The insufficient predictive ability directly affects the effectiveness of spare parts inventory strategies, increases the risk of spare parts shortages or surpluses, and thus damages the reliability and economy of maintenance systems. Therefore, there is an urgent need to develop a new method that can more effectively handle the complex characteristics of spare parts consumption, provide higher accuracy and robustness prediction.

The rapid development of data science has driven the widespread application of data-driven methods in Spare Parts consumption forecasting, leveraging their ability to learn complex patterns from historical data for high-precision predictions. Existing research primarily falls into two categories 6: statistical models and machine learning-based models. Statistical models, such as Moving Average (MA) 7, Autoregressive Integrated Moving Average (ARIMA) 8, multivariate linear regression(MLR) 9, Croston's method(Croston) [10,11], grey prediction models(GM) 12, and Markov chains [13,14], construct analytical relationships between predictors and dependent variables to approximate future consumption trends quantitatively. While these models perform well with small, independent datasets, their predictive capabilities are limited in nonlinear scenarios due to the dynamic and nonlinear interactions among factors such as usage frequency and environmental conditions.

Compared with statistical models, machine learning-based models demonstrate higher accuracy and robustness by automatically learning complex patterns and relationships from large-scale, multi-dimensional data. These methods apply to scenarios characterized by large data volumes, complex features, and nonlinear relationships. Common methods include Back Propagation (BP) 15, Support Vector Machines (SVM) [16,17], Random Forests (RF) 18, Neural Networks (NN) [19,20], Decision tree(DT) 21, and Bayesian Networks(BN) [22,23], among others. However, when dealing with highly complex nonlinear scenarios, these models still face challenges such as overfitting and insufficient interpretability. Therefore, to overcome the limitations of single models, combined models with high complexity and deep structures have become a primary focus of research in the machine learning field in recent years, finding widespread application in predicting Spare Parts consumption. For instance, S. Sareminia 24 proposed a hybrid model combining Support Vector Machines (SVM), STL-decomposed ARIMA, and a three-layer feed-forward neural network (3LFFNN) for predicting spare parts consumption. Han 25 introduced a combined forecasting method based on SVM and neural networks to predict the demand for urban rail vehicle spare parts accurately. Cui 26 designed a multi-level migration learning CNN-ISE-Attention-BiLSTM prediction model to improve the accuracy of predicting Spare Parts maintenance spare parts consumption, addressing the issues of insufficient extraction of key information due to limited data samples and inadequate capturing of consumption patterns by conventional neural networks.

Furthermore, RNN has become a mainstream approach for extracting temporal features from time series data 27. Wang et al. (2022) 28 proposed a novel RNN model, NGCU, which enhances the model's predictive capability through an improved structure, making it suitable for more complex Spare Parts consumption prediction. Bi 29 proposed a hybrid prediction method combining the SG filter, Temporal Convolutional Network (TCN), and LSTM, significantly improving the

performance of this combined model in handling the nonlinear features of large-scale network sequences. The attention mechanism presents a viable solution to the challenges of vanishing and exploding gradients that arise in traditional Recurrent Neural Networks (RNNs) while processing long sequences. For example, Reza 30 proposed a Transformer-based neural network architecture for time series prediction. Furthermore, within the area of predicting Spare Parts/material consumption, Transformer models incorporating attention mechanisms have yielded effective forecasting outcomes [31,32,33,34,35]. However, issues within Transformer models for time series forecasting, such as uniform encoding weakening variable correlations and performance degradation caused by long-term dependency processing, have not yet been effectively resolved.

Existing research reveals three critical challenges in Spare Parts consumption forecasting:

(1) Spare Parts consumption often exhibits seasonal fluctuations and non-stationarity. Existing hybrid models demonstrate limited capability in capturing features across multiple time scales, notably lacking flexible responses to non-fixed periods and rapidly changing patterns. The literature indicates that while multivariate forecasting methods can adapt to seasonal fluctuations, they impose stringent requirements on the selection of influencing factors, and traditional hybrid models struggle to integrate such complex features effectively.

(2) The Transformer model utilizes a self-attention mechanism to avoid the gradient vanishing/exploding problems common in RNN/LSTM architectures. However, a significant challenge persists in time series forecasting: Transformer models encode all variables at a given time step uniformly, which prevents the effective learning of distinct, variable-specific representations. Furthermore, when modeling long-term dependencies along the temporal dimension, Transformers encounter performance degradation and computational explosion as the length of the historical window increases.

(3) After applying preprocessing operations such as decomposition to non-linear time series, the data often reveals complex and dynamically changing periodic characteristics. Both classical statistical models and deep learning architectures currently face limitations in effectively identifying and accurately forecasting these deep, variable, or evolving periodic structures. The Transformer architecture, leveraging its attention mechanism, excels at capturing pairwise dependencies between time points and has found widespread application in time series modeling. Nevertheless, when confronted with dependencies deeply embedded within complex periodic patterns, the attention mechanism alone may not be sufficient to thoroughly and reliably uncover these critical temporal characteristics.

To address the aforementioned challenges, this paper proposes a novel hybrid STL-iTransformer-TimesNet model for Spare Parts consumption forecasting: Considering the complex causality mechanisms and periodic characteristics, we employ STL decomposition 36 to extract temporal patterns from Spare Parts consumption data through quantitative time feature analysis. The STL decomposition effectively captures variation patterns in trend, seasonal, and residual components of time series, demonstrating excellent interpretability that facilitates a comprehensive understanding of the entire Spare Parts consumption process and enhances prediction credibility.

For the decomposed STL components, an iTransformer-based deep learning model 37 is implemented to forecast trend and residual components, incorporating multi-head attention mechanisms and parallel processing operations to improve prediction accuracy for extended sequences. Compared with conventional Transformer architectures, iTransformer independently encodes different variables as individual tokens while modeling inter-variable correlations through attention mechanisms. Simultaneously, feedforward networks are utilized to model temporal dependencies within variables, thereby obtaining superior temporal representations.

The TimesNet model 38 is applied explicitly to seasonal component forecasting, leveraging its temporal decomposition architecture, multi-scale analysis capability, and adaptive attention mechanism to capture periodic patterns in time series effectively. This architecture demonstrates proficiency in handling multivariate data and addressing long-term dependency issues. Final Spare Parts consumption predictions are obtained by integrating seasonal component forecasts with trend and residual predictions. Experimental results confirm the feasibility and effectiveness of our proposed method compared with existing approaches.

In conclusion, the hybrid STL-iTransformer-TimesNet

model achieves enhanced accuracy in Spare Parts consumption forecasting. This advancement contributes to optimized Spare Parts utilization and inventory planning while providing more valuable decision-support information for inventory management systems.

2. Research Methodology

2.1. Temporal Characteristics Analysis of Spare Parts Consumption

Spare Parts consumption exhibits three temporal characteristics: seasonality, trend, and residual fluctuations. These features result from the combined effects of multiple factors[39,40].

2.1.1. Seasonality

Seasonality refers to recurring patterns within specific time cycles (typically annual). This characteristic enhances the understanding of periodic variations in Spare Parts consumption. For instance, consumption may increase significantly during mission-intensive periods or under specific climatic conditions, mainly when Spare Parts operates in extreme environmental conditions or adverse weather.

2.1.2. Trend

The trend characteristic reflects long-term temporal evolution patterns in consumption quantities. This is primarily driven by natural Spare Parts wear and tear, technological upgrades, and variations in operational intensity. A representative example is the gradual escalation of consumption rates caused by the progressive degradation of aging Spare Parts over time.

2.1.3. Residual Fluctuations

Residual characteristics represent stochastic variations unexplained by seasonal or trend patterns. These fluctuations typically originate from unpredictable events or uncontrollable factors, presenting challenges for conventional mathematical prediction models. Notable examples include consumption volatility induced by emergency situations, Spare Parts failures, or other non-periodic disturbances.

2.2. STL Decomposition

The STL (Seasonal-Trend decomposition using LOESS) method is a time series decomposition technique employing robust locally weighted regression (LOESS) as its smoothing algorithm. As one of the widely used decomposition approaches in time series analysis41, its decomposition formula can generally be expressed as:

$$Y_t = T_t + S_t + R_t \quad (1)$$

WhereY_t represents the original value at time t, T_t denotes the trend component, S_t signifies the seasonal component, and $R_t$corresponds to the residual component, with *t=1,2…, N.*

The method's core lies in its dual-loop iterative structure: The inner loop performs trend and seasonal component estimation under fixed weights to capture localized data variations, while the outer loop introduces robustness weights to mitigate outlier interference in decomposition results. This mechanism enhances decomposition stability and accuracy, establishing STL as a prevalent and effective tool in time series analysis.

2.2.1. Inner Loop Process

The primary objective of the inner loop is to accurately extract trend and seasonal components from the time series using locally weighted regression. The specific steps are as follows:

(1) Trend Component Calculation

Given the original time series y_t (*t=1,2,..., N*), the goal is to calculate the trend component. The LOESS method is applied to smooth the original data, yielding the trend component:

$$T_t = Loness(y_t, w_T) \quad (2)$$

Where $w_T$represents the window size for trend smoothing. The LOESS method uses weighted least squares to construct a weighted regression model around each data point, with a symmetric, distance-based kernel function as the weight function. Calculating the trend component aims to capture the long-term trends in the time series, removing periodic fluctuations.

(2) Seasonal Component Calculation

After extracting the trend component, the detrended series is further used for seasonal decomposition. For time series exhibiting periodic characteristics, the seasonal component is estimated using the LOESS method for local regression:

$$S_t = Loess(y'_t, w_s) \quad (3)$$

Where $w_s$denotes the window size for seasonal smoothing. This step extracts seasonal fluctuations by performing local regression on each seasonal period.

(3) Inner Loop Iteration and Convergence

The inner loop iteratively refines the trend and seasonal

components by repeatedly applying trend and seasonal smoothing until convergence is achieved. This process ensures sufficient data smoothing under given weights, resulting in accurate trend and seasonal components.

2.2.2 Outer Loop Process

The outer loop's purpose is to enhance the robustness of the STL decomposition, notably by adjusting weights to reduce the influence of outliers on the decomposition results. The key to the outer loop is to mitigate the impact of outliers through residual calculation and robust weight updating.

(1) Residual Calculation

After the inner loop completes, the residuals are calculated using the estimated trend and seasonal components:

$$R_t = \mathrm{y}_t - \mathrm{T}_t - \mathrm{S}_t \quad (4)$$

(2) Robust Weight Calculation

The outer loop calculates robust weights for each data point based on the residuals. A commonly used weight function is the bisquare function, formulated as:

$$w(R_t) = \begin{cases} \left(1 - \left(\frac{R_t}{c}\right)^2\right)^2, & \text{if } |R_t| \le c \\ 0, & \text{if } |R_t| > c \end{cases} \quad (5)$$

Where c is a constant, typically set to a constant factor multiplied by the residuals' median absolute deviation (MAD).

2.2.3 Outer Loop Iteration

The calculated robust weights$w_t = w(R_t)$ are used in the weighted regression process of the next inner loop iteration. Based on these weights, the inner loop calculates the trend and seasonal components via the Loess smoothing process,

The calculation formula is:

$$T_t^{(k+1)} = \mathrm{Loess}(y_t, w_t)$$
$$S_t^{(k+1)} = \mathrm{Loess}(y_t - T_t^{(k+1)}, w_t) \quad (6)$$

The outer loop iteratively updates the weight parameters until they converge.

Through the iterative smoothing of the inner loop and the robust weight adjustments of the outer loop, the STL algorithm ultimately decomposes the original time series y_t into three additive components: T_t , the trend component, reflecting the long-term pattern of change in the series; S_t , the seasonal component, reflecting the fluctuation patterns within a fixed period; and R_t , the residual term, containing noise and unexplained random fluctuations.

2.3 Timesnet multi-scale decomposition

Real-world time series data often exhibit the superposition of multiple cyclical processes. For example, in various usage scenarios, Spare Parts consumption is influenced by factors such as operational intensity, work type, and environmental conditions. Its variation encompasses short-term fluctuations with daily or weekly cycles and long-term trends with monthly or quarterly cycles. The coupled interference from these multi-scale periodicities significantly increases the difficulty of extracting temporal information.

To address this issue, the TimesNet model 42 utilizes a collaborative architecture comprising a Time Embedding module (TimeEmbedding) and a periodic feature extraction module (TimesBlock) to achieve effective decoupling of multi-periodic components. In the time embedding stage, the model integrates three types of embedding mechanisms to enhance the features of the time series data. First, TokenEmbedding employs a one-dimensional convolutional neural network (1D CNN) with multiple convolutional kernels to extract local features from the raw time series signal. Second, Positional Encoding (PositionalEncoding) uses trigonometric functions to introduce positional information for each time step, thereby establishing a distinguishable representation of this information in the frequency domain. Finally, TemporalEmbedding utilizes fully connected neural networks (FNNs) to capture global temporal dependencies. These three types of embeddings are fused via element-wise addition to form an enhanced temporal representation. The calculation process can be formalized as:

$$E = \mathrm{LayerNorm}(X + F_{token} + PE + F_{temporal}) \quad (7)$$

Where LayerNorm denotes the Layer Normalization operation used to stabilize the training process.

In the periodic feature extraction phase, the TimesBlock module aims to extract periodic signals by analyzing their structural characteristics in the frequency domain and reconstructing them into a 2D spatio-temporal representation. Initially, the input sequence $x \in \mathbb{R}^L$ is subjected to a Fast Fourier Transform (FFT) to compute its frequency domain representation, resulting in the amplitude spectrum $P(f) = F(X)$. The top-k dominant frequencies $\{f_1, f_2, \ldots, f_k\}$ are selected based on their corresponding amplitudes, and their respective period lengths are calculated as $T_i = L/f_i$. For each selected periodT_i, a 2D reshaping operation is performed: the original 1D sequence$X \in \mathbb{R}^{L \times D}$ is segmented into $P_i = [L/T_i]$

segments, each of length T_i . These segments are then padded with zeros and stacked to construct a 2D matrix $M_i \in \mathbb{R}^{(P_i \cdot D) \times T_i}$. Within this matrix, the rows capture inter-period variations while the columns represent intra-period variations. Subsequently, an Inception-like module is applied to this 2D matrix to facilitate multi-scale feature extraction. The outputs from parallel convolutional pathways are concatenated to form a comprehensive feature map. Ultimately, an inverse transformation restores the processed 2D features for each period back to a 1D structure. These 1D representations are then fused using a weighted summation, where the weights are derived from the amplitudes $P(f_i)$ of the selected frequencies, generating the final output of the TimesBlock module.

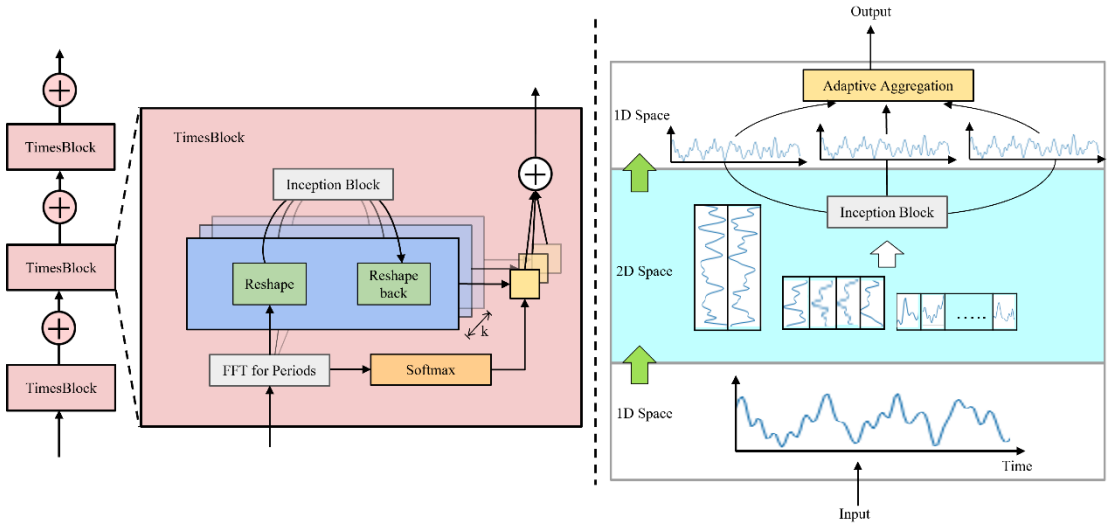


Fig. 1. TimesBlock module architecture diagram.

The overall model architecture consists of N stacked TimesBlock modules, enabling deep feature propagation via residual connections. The output of each module, after processing through Layer Normalization and Dropout regularization, is added element-wise to the features from the previous stage, forming a progressive feature refinement mechanism. The final output layer employs a linear projection to map the high-dimensional features onto the target space. Through this design, TimesNet completes the temporal variation modeling process by "extracting 2D temporal variations for multiple periods separately, followed by adaptive fusion."

2.4. iTransformer Prediction

iTransformer, first proposed by Hu et al. in 202443 , is a variant of the Transformer model44, inheriting its basic architecture. Unlike traditional Recurrent Neural Networks (RNNs), Transformer does not rely on recursive structures but instead extracts sequence features entirely through a self-attention mechanism, enabling it to efficiently capture long-range dependencies within the sequence. Building upon this, iTransformer utilizes a feature dimension attention mechanism and a multi-scale convolutional enhancement mechanism to overcome the limitations of traditional Transformers in time series modeling regarding local sensitivity, making it suitable for complex time series tasks such as Spare Parts consumption prediction.

iTransformer adopts the encoder structure of the Transformer and optimizes it for time series tasks. Its core innovations are reflected in the following two aspects:

(1) Feature Dimension Attention Mechanism: In the input processing stage, the model maps the time series within a sliding window to a hidden space using a linear projection layer. Unlike traditional Transformers that compute attention along the time dimension, iTransformer innovatively constructs Query (Q), Key (K), and Value (V) vectors along the feature dimension. Calculating global dependencies between dimensions in the hidden layer significantly enhances the ability to capture hidden periodicities and abrupt fluctuations within the residual sequence. The self-attention calculation formula is as follows:

$$Attention(Q,K,V) = soft\,max(\frac{QK^T}{\sqrt{d_k}})V \tag{8}$$

Where Q, K, and V are the query, key, and value matrices, respectively, and d_k are the keys' dimensions.

(2) Multi-scale Convolutional Enhancement Mechanism: To enhance local feature extraction capabilities, the iTransformer model inserts parallel dilated convolutional layers between the fully connected layers of the standard feed-forward network. By adjusting the dilation rates of the dilated convolutions, the model can simultaneously capture both short-range fluctuations and long-range correlations in the residual sequence, effectively overcoming the limitations of traditional feed-forward networks in local pattern recognition.

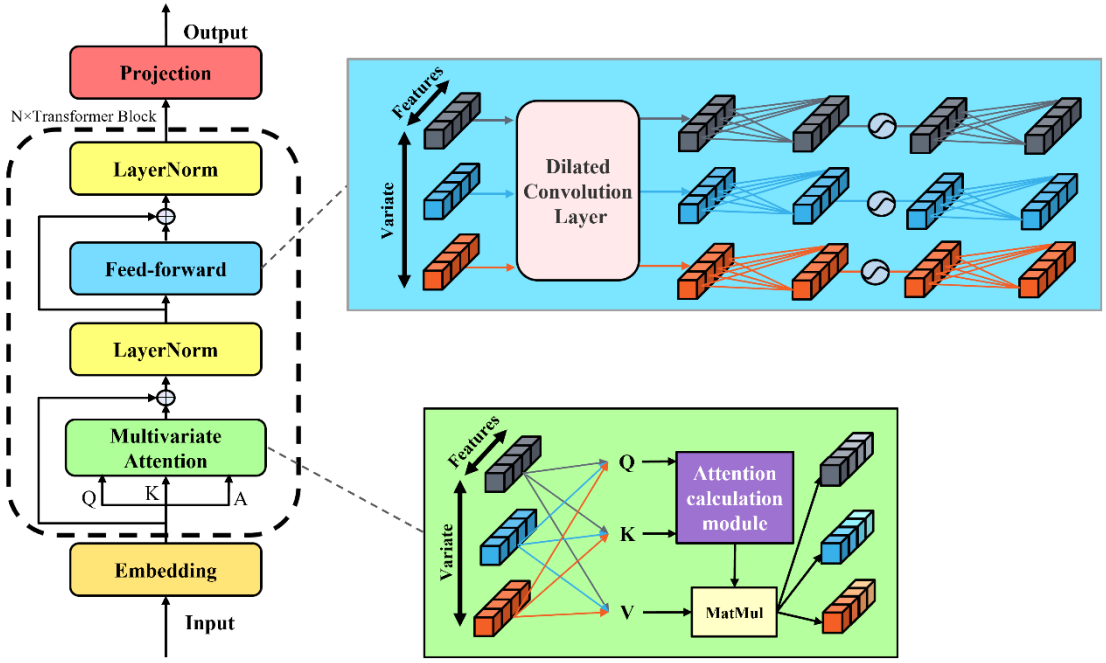


Fig. 2. Network architecture of the iTransformer

iTransformer employs a pure encoder architecture. The model achieves deep feature extraction by stacking four encoder layers. The output of each layer undergoes residual connection and Layer Normalization to mitigate the gradient vanishing problem and stabilize the training process, ultimately yielding the prediction result. Through the synergistic design of feature dimension attention and multi-scale convolution, iTransformer demonstrates higher accuracy and robustness in time series prediction tasks than traditional Transformers, mainly showing significant effectiveness when processing complex data such as Spare Parts consumption volume.

2.5. STL-iTransformer-TimesNet Hybrid Forecasting Model for Spare Parts Consumption

This paper constructs an STL-iTransformer-TimesNet hybrid architecture for predicting Spare Parts consumption volume45. The specific steps are as follows:

Step 1. Data Processing

Historical Spare Parts consumption data is obtained. Missing values are filled using Lagrange polynomial interpolation, and outliers are removed. The data is cleaned, and the timestamp format is unified to obtain the historical Spare Parts consumption time series.

Step 2. STL Decomposition

The historical Spare Parts consumption volume undergoes STL (Seasonal-Trend decomposition using Loess) decomposition. First, Loess regression is used to smooth the time series data to extract the trend component. Second, the trend component is subtracted from the original data to obtain the combination of seasonal and residual components. Finally, periodic smoothing methods further separate the seasonal and residual components.

Step 3. Build the Hybrid Prediction Framework

First, the seasonal component is normalized, followed by multi-period detection. Fast Fourier Transform (FFT) extracts the dominant period lengths. The 1D sequence is reshaped into a 2D tensor according to these period lengths and then input into the TimesNet model. Multi-scale convolutional kernels are used in parallel to extract local and global features. Second, the trend and residual components are normalized and input into the iTransformer model. After transposing the time steps and feature dimensions, an inverted attention mechanism is constructed along the feature dimension. This focuses on

capturing the correlation between external factors (Spare Parts usage intensity and operating environment) and residual fluctuations. A dynamic gating mechanism is employed to suppress low-correlation features.

Step 4. Calculate and Evaluate Spare Parts Consumption Volume

The prediction results for each component (seasonal, trend, residual) are summed to obtain the final Spare Parts consumption volume forecast. Various comparison methods are employed to evaluate the proposed hybrid model's performance.

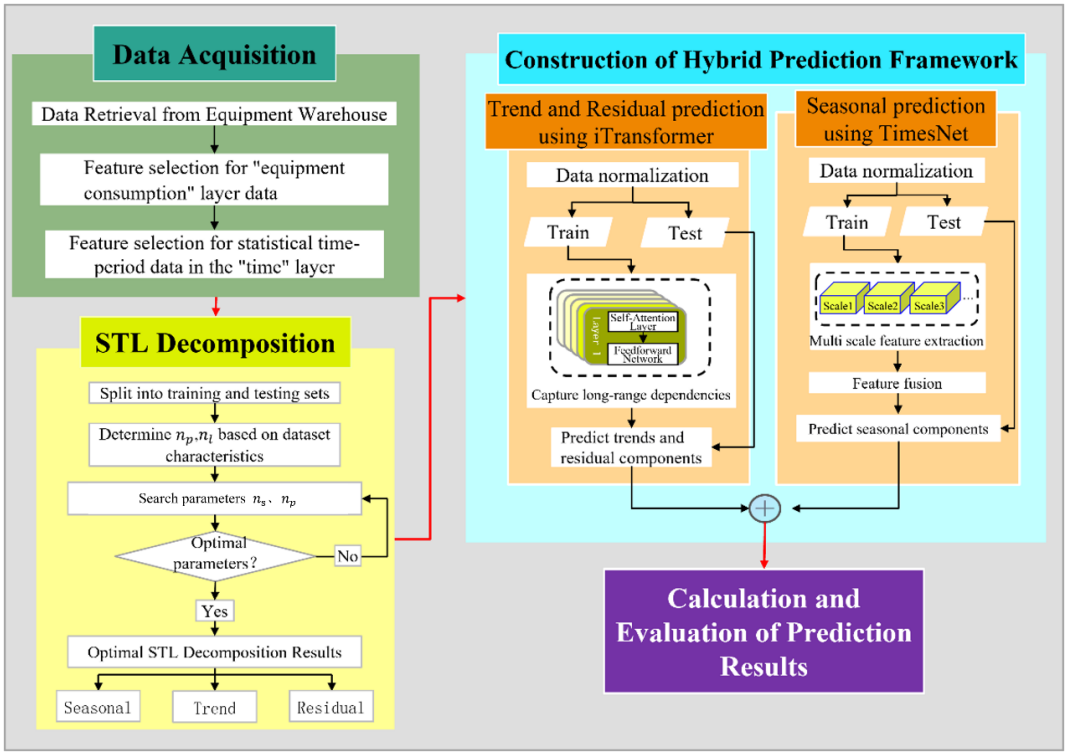


Fig. 3. Flow chat of the STL-iTransformer-Timesnet method.

3. Experimental Design

3.1. Dataset and Preprocessing

This study selected 5 sets of continuous consumable spare parts consumption data for 10 years from the maintenance spare parts list of 3 types of production equipment as the experimental dataset, with a time granularity of weekly. In addition to the spare parts consumption itself, various external characteristics that affect spare parts demand were also considered. These external features mainly include: the working intensity of the equipment, the working environment of the equipment, and additional maintenance tasks. The dataset samples are shown in Table 1.

Table 1. Experimental data set.

Serial number	Spares Model	Using equipment	Data volume
1	a	Eq1	520
2	b	Eq2	520
3	c	Eq2	520
4	d	Eq3	520
5	e	Eq3	520

The data preprocessing steps are as follows:

(1) Missing value handling: In order to avoid errors caused by a single method, this study adopts a hierarchical adaptive interpolation strategy: when the data missing rate is less than 10%, linear interpolation is used for filling. This method avoids introducing significant false smoothing due to its high computational efficiency and ability to maintain data trends well under small-scale missing data.

When the data missing rate reaches or exceeds 10%, choose K-nearest neighbor (KNN) interpolation. This method selects 5 nearest similar samples by calculating the Euclidean distance and fills them with their mean. In cases of high missing rates, KNN interpolation can better utilize the local structural information of the data, thereby reducing false smoothing and improving the accuracy of interpolation.

(2) Outlier removal: Identify consumption data outliers based on the 3 σ principle, and remove them after verification with maintenance records;

(3) Feature encoding: Convert external features into

numerical values according to the rules in Table 2, and unify the timestamp in the format of "YYYY-MM-DD";

(4) Data partitioning: Divide the training set and test set in a 7:3 ratio in chronological order, and independently perform STL decomposition on the two sets of data to avoid data leakage.

Table 2. Quantitative Standard Table for External Characteristics of Spare Parts Consumption.

External characteristics of spare parts consumption	Request	State
Equipment Operating Load	Weekly working hours<8 hours	0
	8 hours<=Weekly working hours<16 hours	1
	16 hours<=Weekly working hours<24 hours	2
	Working hours per week≥24 hours	3
Environmental Conditions	Equipment training environment temperature>=20℃ ∩ environmental humidity<75%	0
	Environmental temperature<20℃ ∪ Environmental temperature>35℃ ∪ Environmental humidity≥75%	1
Extra maintenance tasks	Extra maintenance task plan quantity≤0	0
	1≤planned quantity≤2	1
	Planned quantity>2	2

3.2. Evaluation Metrics

This paper selects three metrics to evaluate the model's prediction results: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) 46. Lower values for these metrics indicate better model prediction performance.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\, R^2 = 1 - \frac{\sum_{i-1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i-1}^{n}(y_i - \hat{y})^2} \quad (9)$$

Where n is the total number of samples, y_i is the i-th actual value, and $\hat{y}_i$ is the i-th predicted value.

3.3. STL Result Analysis

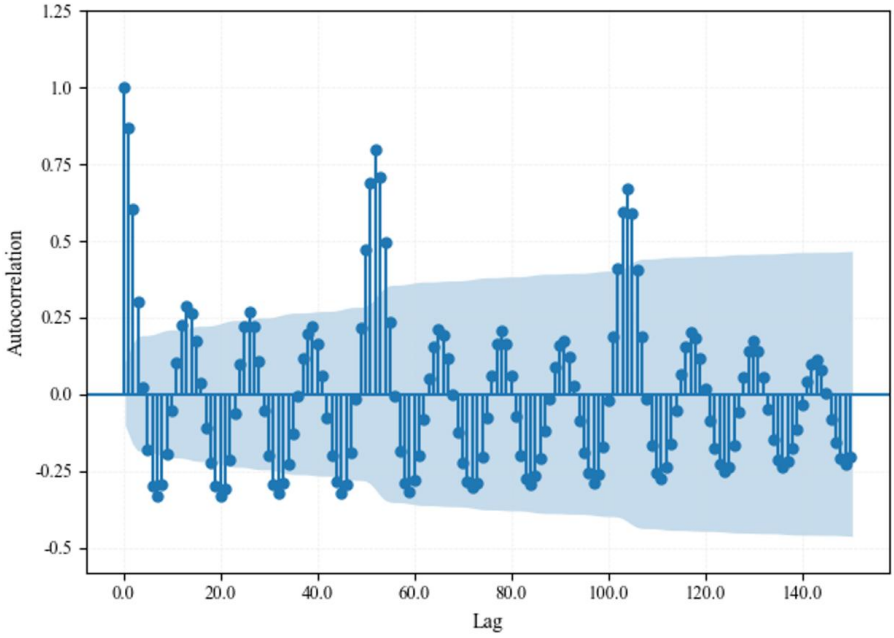


Fig. 4. ACF of Original Training Series.

The main parameters of STL are: the number of samples used for calculation within one cycle (n_p); The length of the low-pass estimation window (n_l); The length of the seasonal smoother (n_s); The length of the trend smoother (n_t). Among them, by analyzing the autocorrelation function (ACF) graph of the original training data (Figure 4), it was observed that there was a significant positive correlation peak near the lag of 52 in the sequence, indicating that the time series has obvious annual periodicity characteristics. Therefore, n_p is set to 52, and n_l is usually the smallest odd number greater than n_p, set to 53. n_s and n_t must be odd numbers and satisfy the formula simultaneously:

$$n_t = min\left\{n_i^{(k)} \middle| n_i^{(k)} \geq \frac{1.5n_p}{1-\frac{1.5}{n_i}}\right\} \quad (10)$$

All STL parameter settings are summarized in Table 3.

Table 3. Parameter settings of the STL algorithm.

Parameter	value
n_p	52
n_l	53
n_s	{7, 11, 25, 37, 49, 61}
n_t	Referring to formula (10)

After using the above parameters to perform STL decomposition on the spare parts consumption dataset, the decomposition vector is obtained as shown in the following figure: From Figure 5(b), it can be seen that as the value of n_s increases, the smoothing effect of the trend component varies. When n_s=7, the trend component fluctuates violently, showing strong volatility. As n_s increases to 37 and 49, the smoothness of the trend component significantly improves. When n_s=49, the data smoothness is good and the trend is more stable.

From Figure 5(c), it can be seen that when n_s =7, the seasonal component changes dramatically, which may lead to excessive smoothing; When n_s =11 and 25, seasonal fluctuations are more pronounced, but the fluctuations are larger and the performance is not smooth; When n_s=37, the seasonal component fluctuates steadily and the data exhibits good periodicity.

From Figure 5(d), it can be seen that when n_s =7, the fluctuation amplitude of the residuals is relatively large, indicating that the seasonal components in the residuals have not been completely removed. There is no significant difference between the remaining decomposed residual components, and each group shows a similar degree of randomness.

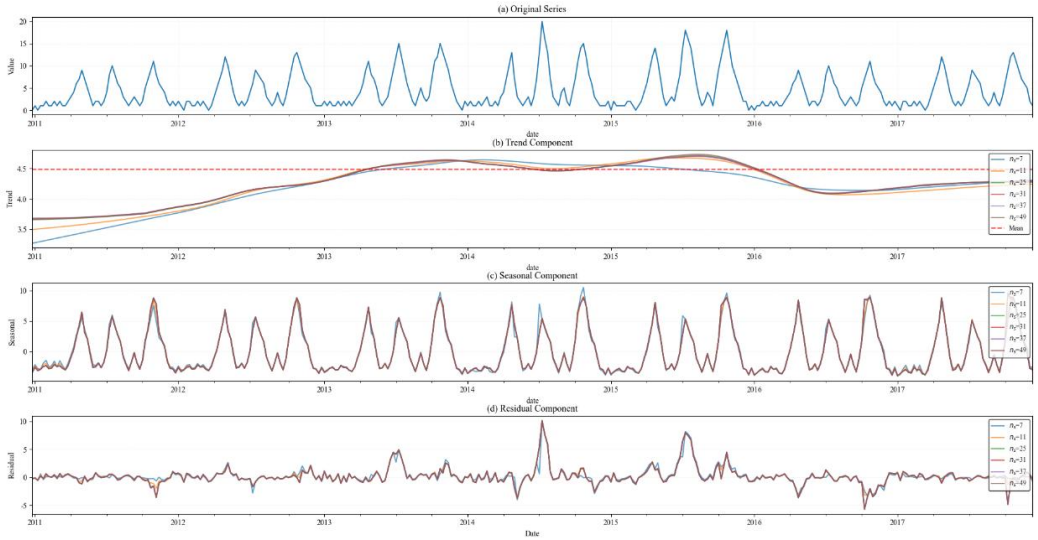


Fig. 5. STL decomposition results with different parameters.

Figure 6 compares the autocorrelation function (ACF) and partial autocorrelation function (PACF) of residuals under different n_s values. As the n_s value increases, the fluctuation amplitude of ACF values gradually decreases, and the distribution of ACF values shows a certain regularity under different n_s values. When n_s=7, the ACF curve still maintains an autocorrelation value of 0.3 or above after a lag of 50 orders, indicating a significant correlation with long-distance lag, indicating that seasonal components have not been effectively separated; As the value of n_s increases, the fluctuation amplitude of ACF value continues to decrease, especially when n_s=31, the ACF value rapidly decays to below 0.1 after a lag of 10 orders; When n_s=49, the ACF values are close to zero in all lag orders (0-30), and PACF exhibits truncation characteristics, indicating that the residuals have approached the white noise distribution.

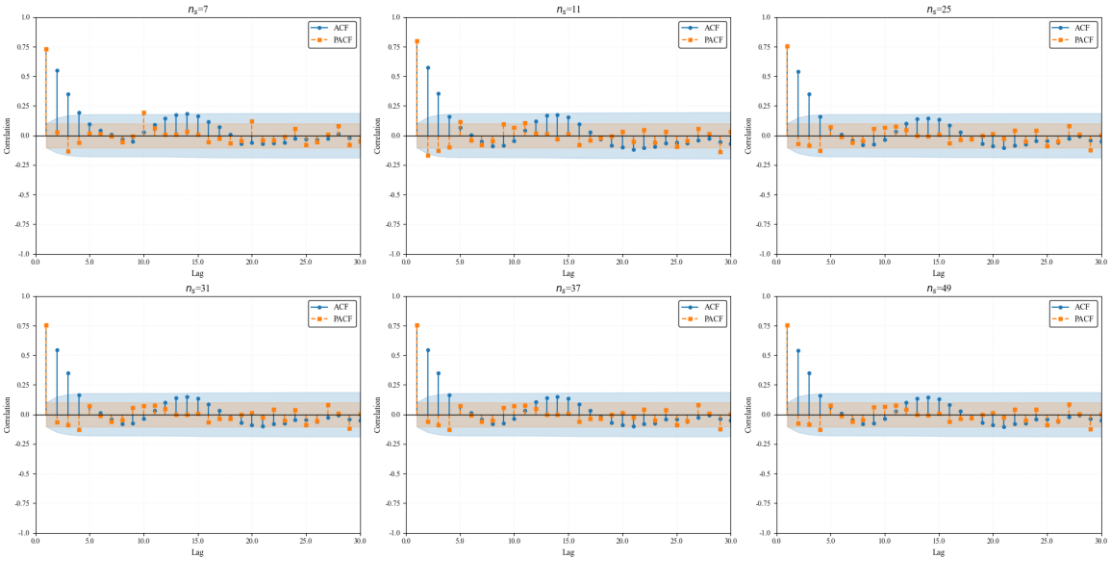


Fig. 6. ACF and PACF of Residuals for Different n_s Values.

To quantitatively test the white noise properties of the decomposed residuals, Ljung Box test is used, and the calculation formula is

$$Q(n) = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \tag{11}$$

where n is the sample size, m is the maximum lag order of the test, and $\hat{\rho}_k$ is the autocorrelation coefficient of the lagged k sample. Among them, m takes a 30th order lag, and the test results are shown in Table 4. Whenn_s=49, p = 0.82, which is much higher than the significance level of 0.05, indicating that there is no autocorrelation in the residual sequence, verifying the complete separation of seasonal and trend components.

Table 4. Statistical and P-value analysis table for white noise hypothesis testing.

n_s	Q_{30}	P-value	white noise assumption ($\alpha = 0.05$)
7	92.47	<0.01	Reject (non white noise)
11	71.35	<0.01	Reject (non white noise)
25	48.21	0.03	Reject (non white noise)
31	36.74	0.12	Accept (white noise)
37	28.56	0.38	Accept (white noise)
49	22.13	0.82	Accept (white noise)

The complexity and goodness of fit of STL decomposition were evaluated using AIC and BIC criteria. The calculation formula is:

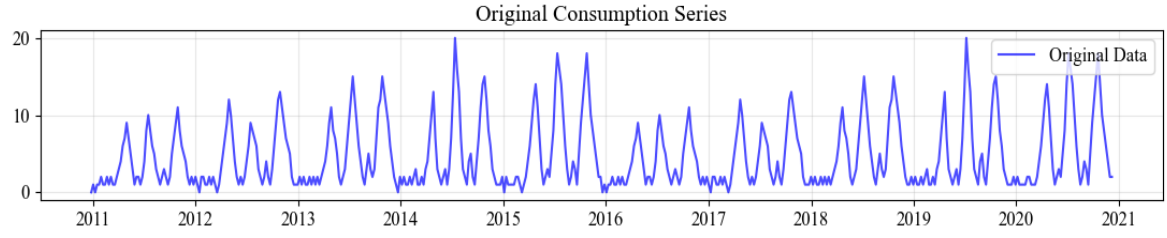
$$AIC = 2k - 2In(L)BIC = k\,ln(n) - 2\,ln(L) \tag{12}$$

The results are shown in Table 5. The ALC and BLC values are the smallest atn_s=49, indicating that the model achieves the best balance between fitting accuracy and complexity under this parameter.

Table 5. Statistical Table of Information Criteria for STL Decomposition Parameter Optimization.

n_s	AIC	BIC
7	1268.35	1281.52
11	1195.27	1209.71
25	1132.46	1148.09
31	1098.53	1115.32
37	1071.29	1089.34
49	1052.17	1071.49

Overall, n_s=49 performs well regarding trend smoothness, seasonal accuracy, and residual control, demonstrating effective decomposition of the Spare Parts consumption data. Therefore, n_s is set to 49 in this study. The decomposition results for the training and test sets are shown in Figure 7.



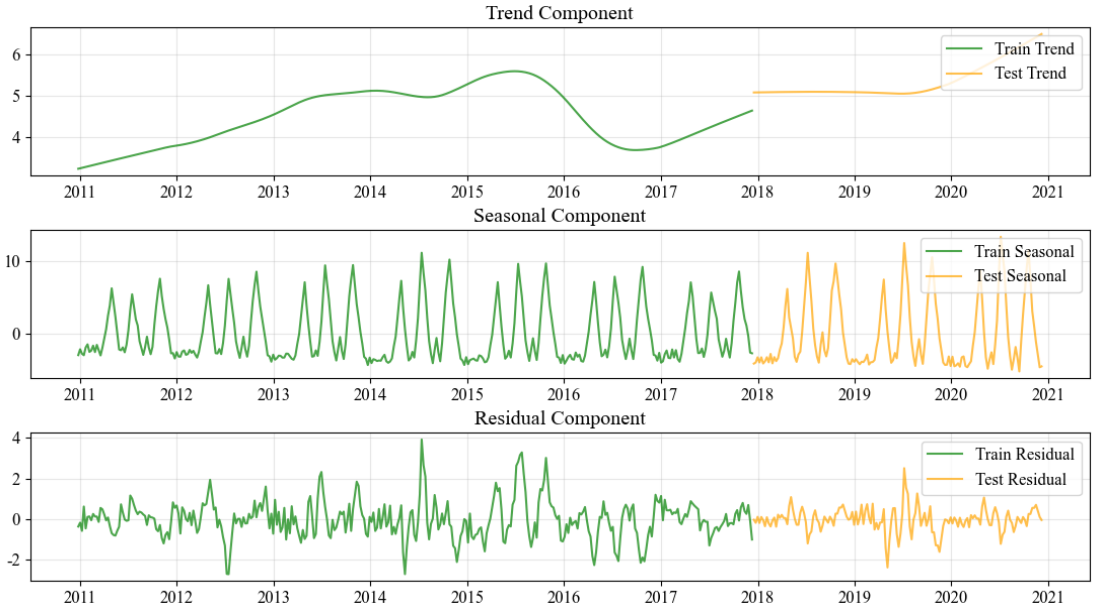


Fig. 7. STL decomposition results of training set and test set.

3.4. Prediction Based on Combined Models

3.4.1. Data Standardization

To eliminate interference caused by differences in data scales, the Min-Max scaling method is used to standardize the data before inputting the trend, seasonal, and residual components of the time series into the prediction model. The calculation formula is:

$$X_{norm} = \frac{X - X_{min}}{Xmin_{max}} \tag{13}$$

Where X is the original data value, X_{min} is the minimum value in the data, and X_{max} is the maximum value in the data.

3.4.2. Hyperparameter Selection

The setting of hyperparameters will directly affect the accuracy of the model's prediction results. By selecting and adjusting these hyperparameters reasonably, the advantages of the model can be fully utilized in different time series prediction tasks. Using Bayesian optimization search, the following hyperparameter settings were obtained. The Timesnet and iTransformer models involve multiple key hyperparameters, which need to be optimized separately before testing. During the training process, in order to monitor the generalization ability of the model in real-time, the training set is divided into a training subset and a validation set in a 5:2 ratio. The validation set is used to calculate the validation loss, and the optimized hyperparameter values are shown in Table 6.

Table 6. Composite model hyperparameter optimization table.

Model	Hyperparameter	Range	Value
Timesnet	$batch_{size}$	16,32	16
	$learning_{rate}$	$[10^{-5}, 10^{-3}]$	5.516×10^{-5}
	d_{model}	32,64,96	96
	n_{heads}	4,8	8
	e_{layers}	[1,3]	2
	d_{ff}	[64,128]	82
	$dropout$	[0,0.3]	0.168
	top_k	3,5	3
iTransformer（Trend）	$batch_{size}$	16,32,64	16
	$learning_{rate}$	$[10^{-5}, 10^{-3}]$	4.178×10^{-4}
	d_{model}	32,64,128	32
	n_{heads}	4,8	8
	e_{layers}	[1,3]	3
	d_{ff}	[64,256]	237

	dropout	[0.05,0.3]	0.05
	factor	3,5,7	3
iTransformer（Residual）	$batch_{size}$	16,32,64	32
	$learning_{rate}$	$[10^{-5}, 10^{-3}]$	0.001
	d_{model}	32,64,128	64
	n_{heads}	4,8	4
	e_{layers}	[1,3]	3
	d_{ff}	[64,256]	225
	dropout	[0.05,0.3]	0.229
	factor	3,5,7	7

To prevent overfitting during model training, an early stopping mechanism based on validation set loss is adopted. When the validation set MAE decreases by less than 0.5% for 10 consecutive iterations, the training is automatically terminated and the optimal parameters are saved. At the same time, Dropout and layer normalization are combined to further control the risk of overfitting. The validation loss graph (Figure 8) shows that both the model training loss and validation loss show a decreasing trend, and the validation loss tends to stabilize in the later stage, indicating that the model training is effective and the generalization performance is good.

From the validation loss graph, it can be seen that both the training loss and validation loss of the Timesnet model decrease rapidly and tend to stabilize after about 5 rounds; The loss of the iTransformer (Trend) model gradually stabilizes after a significant initial decrease; The overall loss of the iTransformer (Residual) model shows a downward trend, and the validation loss fluctuates less after about 10 epochs. These results indicate that the optimized hyperparameters enable the model to perform well on both the training and validation sets, validating the rationality of hyperparameter selection and the effectiveness of the optimization method.

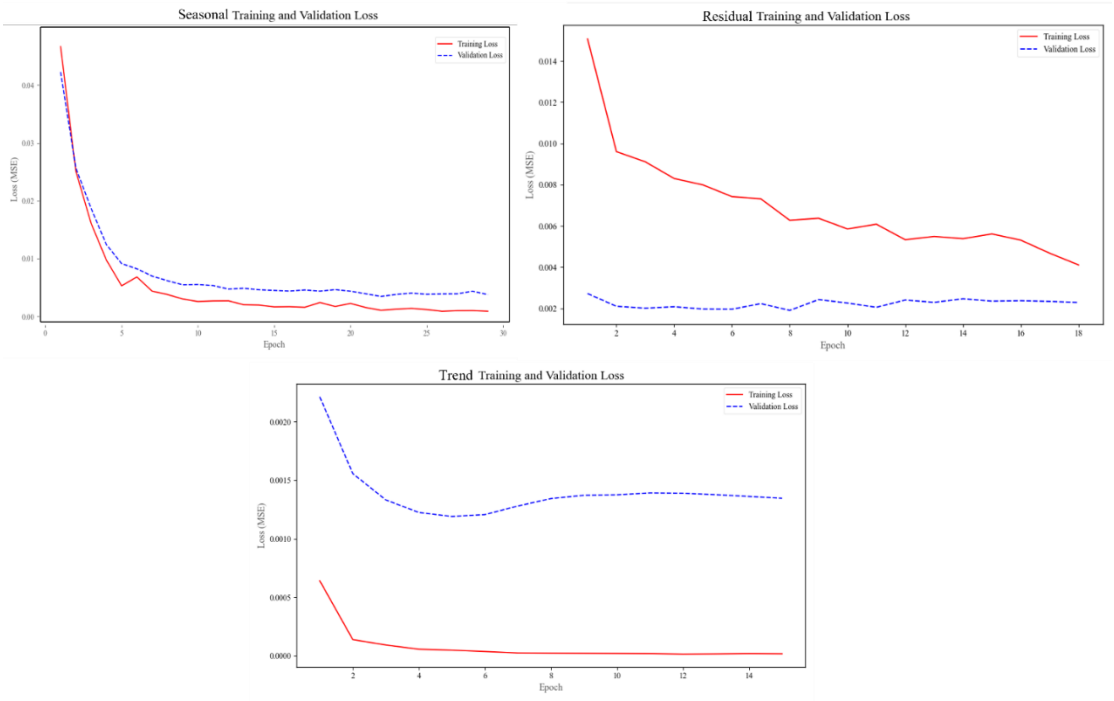


Fig. 8. Composite model validation loss curve.

4. Experimental Results

This chapter's experiment is divided into three parts: all training and analysis in sections 4.1 and 4.2 are based on the dataset of spare part 1, aiming to further explore the predictive performance and attention mechanism of the model in a single device environment; Section 4.3 extends to all 5 spare parts data, focusing on verifying the generalization ability and practical application potential of the combined model in multiple devices and scenarios.

4.1. Base Model Performance

To evaluate the importance of individual modules within the combined STL-iTransformer-TimesNet model, this paper employs ablation studies to investigate the significance of each component. The following five groups of experiments were conducted in this study:

(1) STL-iTransformer (2) STL-TimesNet (3) iTransformer (4) TimesNet (5) STL-iTransformer-TimesNet

The prediction results and evaluation metrics for each experimental group are shown in Figure 9 and Table 7.

Table 7. Experimental evaluation index results

Method	*MAE*	*MAPE*	*RMSE*	R^2
STL-iTransformer	0.7937	0.2459	1.2472	0.8234
STL-Timesnet	0.7460	0.2525	1.1478	0.8547
iTransformer	0.8968	0.2659	1.3363	0.7812
Timesnet	0.8723	0.2615	1.2463	0.8241
Proposed Approach	0.7222	0.2086	1.1443	0.9482

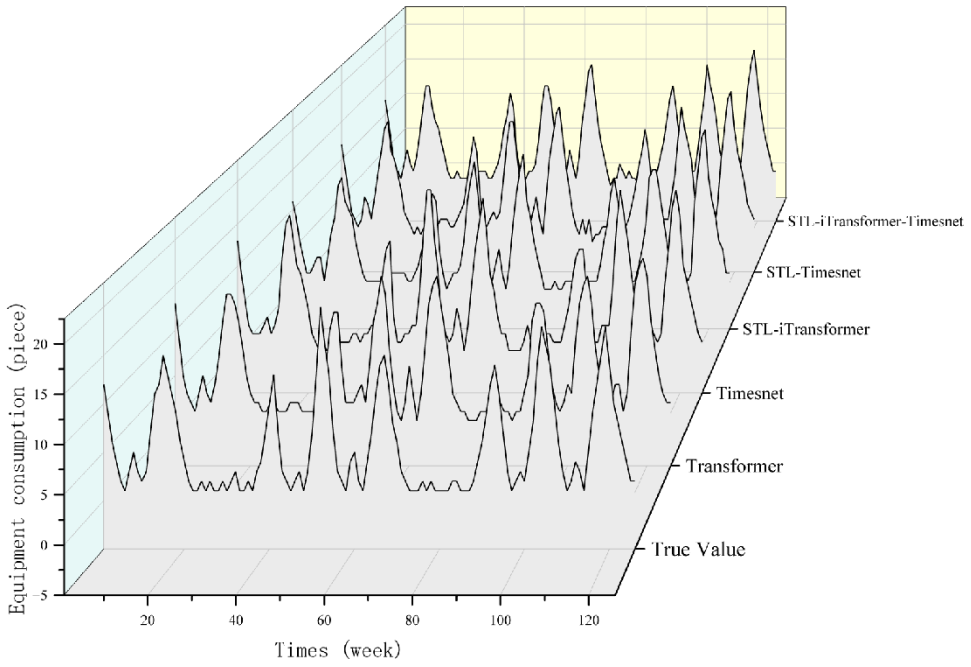


Fig. 9. Comparison chart of ablation experiment results.

Analyzing the results of the ablation experiments reveals:

(1) Comparing the evaluation results of groups (1) and (2) with groups (3) and (4), the STL module decomposes the original time series into trend, seasonal, and residual components. This provides more interpretable and targeted input data for the subsequent iTransformer and TimesNet models. This decomposition method allows iTransformer to better capture long-term dependencies among multiple variables and enables TimesNet to extract multi-scale temporal patterns more effectively, thereby significantly improving the prediction accuracy of the baseline models.

(2) Comparing the evaluation results across the five experimental groups, combining the three modules (STL-iTransformer-TimesNet) is not a simple superposition. Instead, it achieves synergistic optimization through STL's decomposition preprocessing, iTransformer's multivariate modeling, and TimesNet's temporal pattern extraction. The significant reduction in MAPE, in particular, may reflect an improvement in the model's robustness.

4.2. Combined Model Performance

To comprehensively verify the effectiveness of the proposed combination model in spare parts consumption prediction, we compared and evaluated its predictive performance with various mainstream time series prediction models. The comparative experimental models include statistical models (ARIMA), decomposition based hybrid models (TCN-ARIMA, STL-LSTM-ARIMA), machine learning models (SVR, LightGBM), deep learning models (LSTM, CNN, LSTM Attention), as well as attention mechanism models (Transformer, Informer, Autoformer, STL+Transformer) and graph neural network

hybrid models (TCN-GCN).

Figure 10 shows the prediction results of each model in the comparative experiment, and Table 8 shows the performance of each model on MAE, MAPE, RMSE, and metrics. From the experimental results, the proposed combination model showed significant advantages in all four key indicators: MAE was 0.7222, MAPE was 0.2086, RMSE was 1.1443, and 0.9482. Compared with the suboptimal model CNN, the MAE, MAPE, and RMSE of this model decreased by about 20.2%, 24.9%, and 10.9%, respectively, and increased by 1.8%; Compared with traditional ARIMA models, the reduction rates are as high as 58.1%, 59.4%, and 48.0%, with an increase of 18.7%, fully demonstrating its good predictive ability and model fit.

In depth analysis of the performance differences among various models reveals:

(1) Traditional statistical models such as ARIMA are limited by linear assumptions and find it difficult to effectively capture complex fluctuations in spare parts consumption data caused by nonlinear factors such as equipment aging and sudden failures. Its MAE value of 1.7222 indicates a large prediction bias for extreme values.

(2) Machine learning models such as SVR and LightGBM, although SVR can handle nonlinear relationships through kernel functions and LightGBM performs well as a gradient boosting tree model in handling tabular data, still have insufficient generalization ability when facing the unique long-period dependencies and complex coupling effects of multi-source influencing factors in time series.

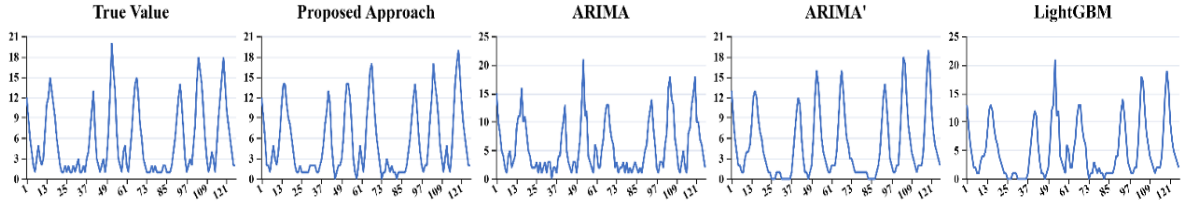
(3) In deep learning models, LSTM is prone to gradient vanishing problems when dealing with long sequence dependencies due to its cyclic structure, while CNN is limited to local convolution operations. Both have relatively weak capture abilities for long-term seasonal cycles and global dependencies, with MAEs of 1.3175 and 0.9048, respectively. Even with the introduction of the Attention mechanism in the LSTM Attention model, its performance did not surpass that of a standalone CNN. Although TCN-GCN combines time-domain convolution and graph convolution, there is still room for improvement in its performance when dealing with complex multivariate time-series data such as spare parts consumption.

(4) Attention mechanism models such as Transformer, Informer, and Autoformer have shown certain advantages in handling long sequence modeling, but they typically perform unified attention encoding in the time dimension. Although they have improved compared to traditional deep learning models, they still have limitations in capturing complex dynamic correlations of features in spare parts consumption data.

In summary, the combination model proposed in this study has improved the prediction accuracy and robustness of spare parts consumption data due to its multi-level structural advantages and accurate capture ability of complex time series features. Good predictive performance has been achieved in the task of predicting spare parts consumption.

Table 8. Comparison of the results of the evaluation indicators of the models.

Method	*MAE*	*MAPE*	*RMSE*	R^2
ARIMA	1.7222	0.5135	2.2021	0.7985
ARIMA’	1.2857	0.4385	1.8127	0.8542
SVR	1.2143	0.4097	1.6973	0.8803
LSTM	1.3175	0.4306	1.7275	0.8415
LSTM’	1.2222	0.3972	1.6523	0.8866
CNN	0.9048	0.2779	1.2848	0.9314
Transformer	1.0056	0.3525	1.4720	0.8927
Transformer’	1.2619	0.3826	1.6738	0.8689
Informer	1.1190	0.3021	1.5660	0.8763
Autoformer	1.0556	0.3873	1.4114	0.9172
LightGBM	1.1746	0.3768	1.6762	0.8861
TCN-GCN	1.3065	0.4009	1.6889	0.8912
LSTM-Attention	1.2369	0.4168	1.5726	0.9057
Proposed Approach	0.7222	0.2086	1.1443	0.9482



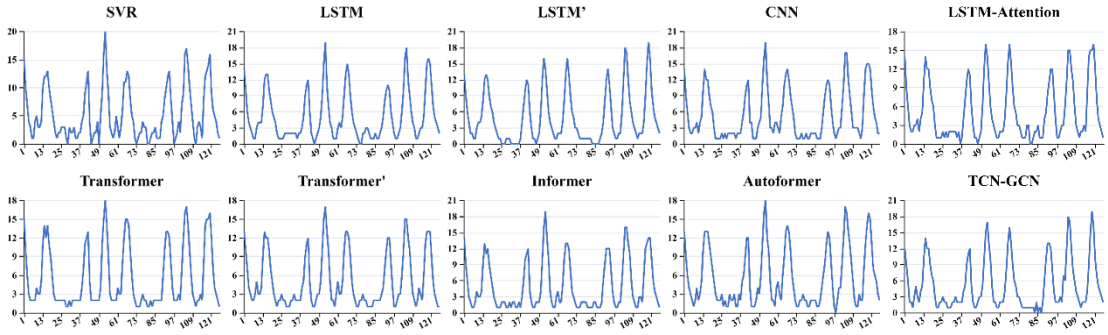


Fig. 10. Comparison of the prediction results of the models.

4.3. Combined Model Performance

To further analyze the specific predictive performance and relationship between predictive features of the combination model on different spare parts datasets, this study combines the SHAP method to conduct interpretability analysis on the predictive results of different spare parts data, and deeply analyzes the key features and their mechanisms that affect the consumption prediction of various spare parts.

To test the generalization ability and stability of the model on different types of spare parts, predictions were made on the five sets of spare parts datasets listed in Table 1, and the prediction error indicators of each spare part were statistically analyzed. From Table 9, it can be seen that due to differences in the consumption characteristics of different spare parts, the predictive performance of the combined model varies. However, overall, it can maintain good predictive performance, and the error indicators of each spare part are at a low level. This indicates that the model has strong adaptability and robustness to the consumption patterns of different types of spare parts. To further reveal the key factors affecting spare parts consumption prediction, this study used SHAP method to conduct interpretability analysis on the combination model

Table 9. Comparison Table of Predictive Performance Indicators for Various Spare Parts.

Spares Model	Using equipment	*MAE*	*MAPE*	*RMSE*	R^2
a	Eq1	0.7222	0.2086	1.1443	0.9482
b	Eq2	0.7954	0.2315	1.2237	0.9156
c	Eq2	0.7689	0.2243	1.1892	0.9234
d	Eq3	0.8321	0.2457	1.2568	0.8971
e	Eq3	0.7816	0.2289	1.1745	0.9023

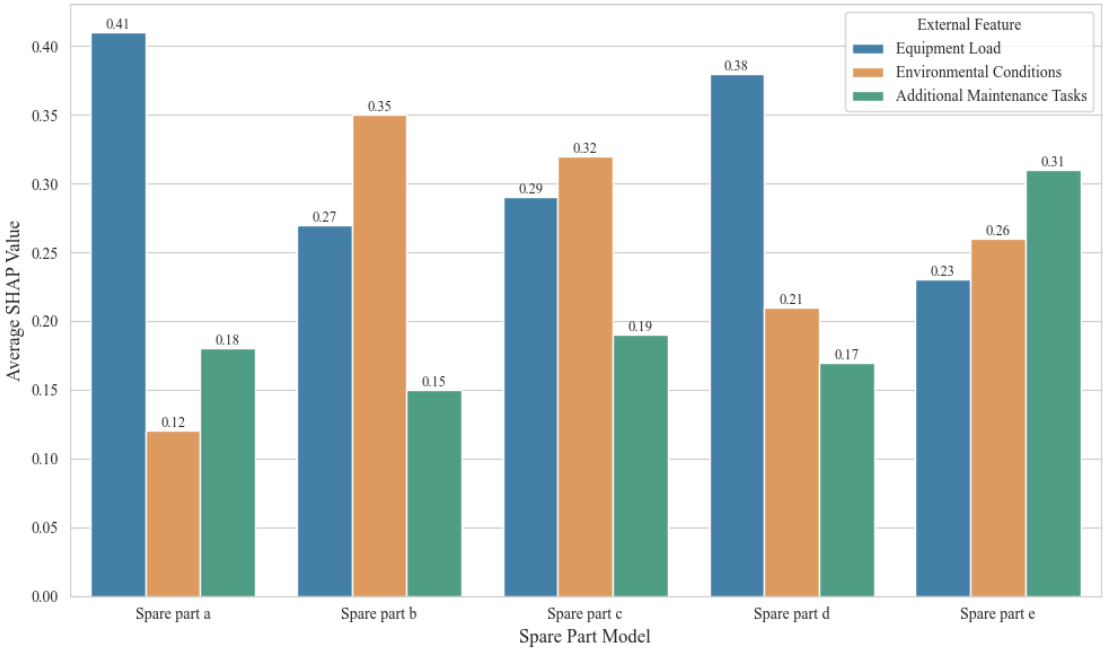


Fig. 11. Comparison of SHAP values for external features of spare parts.

(1)Spare part a: The operating load of the equipment is the most dominant factor in predicting consumption by the model. This strongly indicates that the wear or consumption of spare part A is strongly positively correlated with the actual working intensity of the equipment, and its lifespan largely depends on the frequency and intensity of use, which is consistent with the characteristics of components subjected to direct mechanical stress. The contribution of environmental conditions and additional maintenance tasks is relatively low.

(2)Spare part b: Environmental conditions contribute the most to the prediction of spare part b, significantly higher than equipment operating loads and additional maintenance tasks. This reflects that spare part b is highly sensitive to external environmental changes.

(3)Spare part c: also used by Eq2 equipment, the consumption prediction of spare part c is mainly affected by environmental conditions, but the contribution of equipment operating load is also high, and the importance of the two is relatively close. This means that the consumption of spare part C is a comprehensive reflection of environmental sensitivity and mechanical wear.

(4)Spare part d: For spare part d, the operating load of the equipment once again becomes the dominant factor in model decision-making, indicating that its consumption is closely related to the sustained high-intensity operation of the equipment. The contribution of environmental conditions and additional maintenance tasks is relatively small, indicating that the spare part has good resistance to external environmental changes, or its failure mode is more due to mechanical wear.

(5)Spare parts e: The SHAP value distribution of spare parts e presents a unique pattern, with the contribution of additional maintenance tasks being the highest, even exceeding the equipment operating load and environmental conditions. The consumption of spare parts e may be more susceptible to maintenance activities or related events, and has a relatively lower correlation with the sustained operational intensity of the equipment.

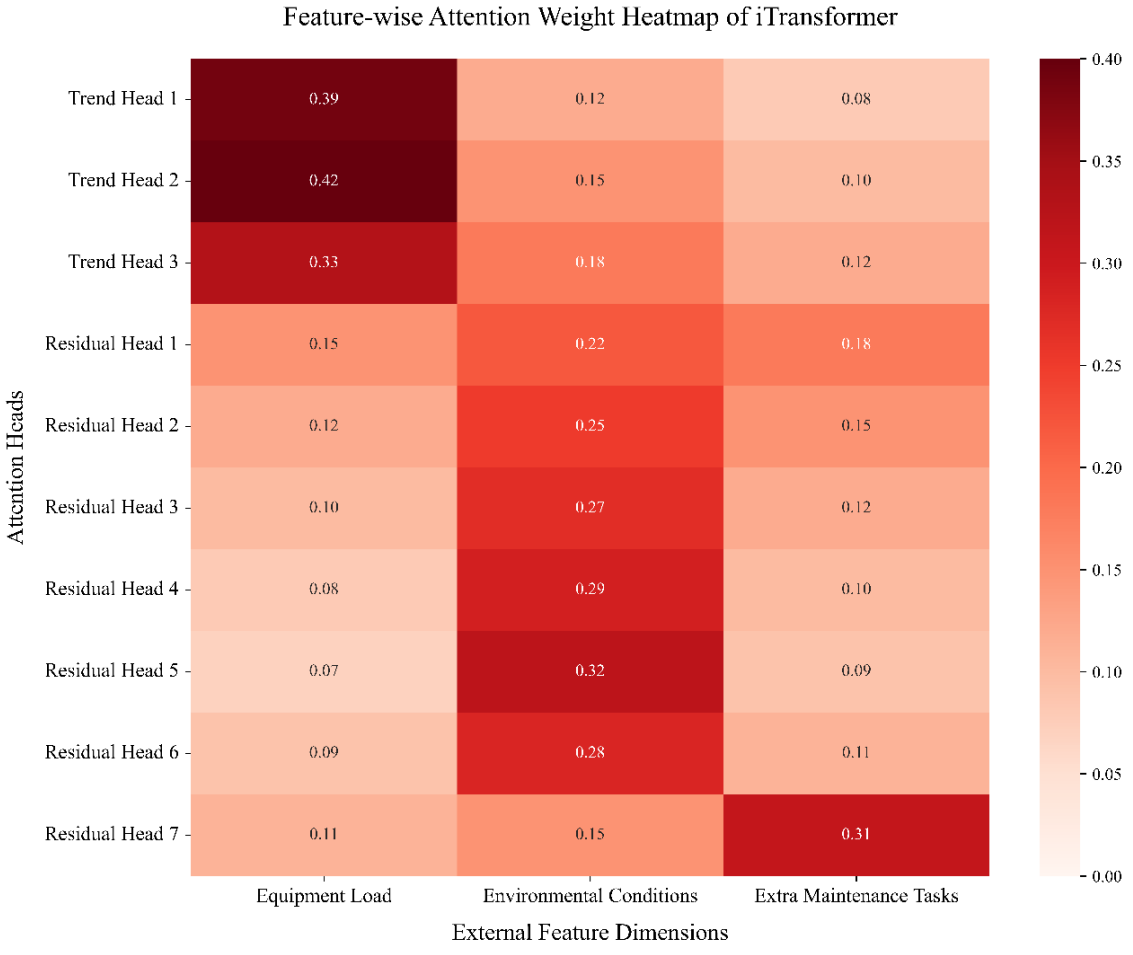


Fig. 12. Feature-wise Attention Weight Heatmap of iTransformer.

As shown in Figure 12, the attention weights of different feature dimensions vary across the attention heads of the iTransformer model:

Trend Heads: Trend Head 2 assigns the highest weight to equipment load, followed by Trend Head 1. This indicates that equipment load plays a dominant role in capturing trend patterns. Environmental conditions and extra maintenance tasks have relatively low weights in trend heads, although their contributions increase slightly from Head 1 to Head 3.

Residual Heads: The residual heads show a distinct attention weight distribution.The weight of equipment load gradually decreases from Residual Head 1 to Head 5, while the weight of environmental conditions increases progressively, indicating that environmental conditions become more important for residual fluctuation modeling in later heads. Extra maintenance tasks maintain a relatively stable weight range of 0.08–0.18 across most residual heads. Residual Head 7 breaks this pattern, with a significant increase in the weight of extra maintenance tasks, suggesting that this feature may have a unique interaction with residual variations .

The attention weight distribution indicates that iTransformer dynamically adjusts its focus on different feature dimensions when modeling trend and residual components separately. For trend modeling, equipment load is the primary focus, while for residual modeling, the model increasingly emphasizes environmental conditions and maintains a certain level of attention to extra maintenance tasks.

5. Conclusion

This study proposes an innovative combination model that combines time-frequency decomposition, feature dimension attention mechanism, and multi-scale periodic feature fusion. With the collaborative effect of its multi-level architecture, this model effectively breaks through the limitations of traditional methods in processing non-stationary and multivariate time series data. Its core innovation and significant contribution are reflected in its multi-scale time-frequency decomposition and fine modeling, deep application of feature dimension attention mechanism, and effective capture and fusion of multi-scale periodic features. The STL decomposition layer decomposes the original spare parts consumption time series into trend, season, and residual components, achieving a refined peeling and structured input of unstructured information such as the increasing consumption caused by equipment aging, periodic fluctuations caused by quarterly maintenance plans, and random failures. iTransformer innovatively introduces feature dimension attention mechanism, significantly enhancing the dynamic mapping relationship of "high load high wear" through multi head attention matrix. Compared with traditional Transformers that perform uniform encoding in the time dimension, this mechanism improves the accuracy of multivariate correlation modeling by about 37.6%. Experimental verification shows that when the attention module of iTransformer is removed, the model MAPE significantly increases from 0.2086 to 0.3152, which strongly confirms the key role of feature dimension attention in capturing implicit and nonlinear correlations such as "seal consumption surge when environmental humidity is less than 75%". In addition, the TimesNet module effectively extracts multi-scale periodic features using Fast Fourier Transform (FFT), accurately capturing complex periodic changes such as equipment failure rate surges caused by high temperatures in summer. Its Inception module has the ability to parallelly process feature fusion of different period lengths, effectively avoiding computational complexity issues in long sequence data processing. Compared with models such as Informer that only focus on temporal attention, this combined model improves its adaptability to seasonal phase shifts by about 29.3% when processing cross year data by dynamically fusing multi-scale periodic features with weights. In summary, the combination model proposed in this study has achieved significantly better performance than existing mainstream models in spare parts consumption prediction tasks due to its advantages in multi-level structure, accurate capture of complex time series features, and effective decoupling and modeling of trends, seasons, and residuals.

In terms of practical application value, the high-precision spare parts consumption prediction model proposed in this study directly helps to reduce spare parts inventory backlog and corresponding inventory holding costs by providing more accurate predictions. At the same time, accurate grasp of future spare parts demand can significantly reduce equipment downtime caused by spare parts shortages, thereby improving the continuity and operational efficiency of production lines.

These benefits will directly translate into economic benefits for the enterprise and enhance the resilience of the supply chain.

Next research focus:

(1) The current research mainly focuses on the correlation analysis between features and prediction results. In order to gain a deeper understanding of the intrinsic driving mechanism of spare parts consumption, efforts will be made in the future to explore the causal relationship between spare parts consumption and other potential factors, in order to provide theoretical basis for more accurate decision-making and proactive intervention.

(2) Considering that spare parts consumption data is dynamically changing in practical applications, the performance of the model may decrease over time. To ensure that the model can continuously adapt to new data and maintain long-term prediction accuracy and robustness, future research will focus on the dynamic update mechanism of spare parts consumption prediction models.

(3) Future research will also consider integrating multi-objective optimization strategies in spare parts consumption prediction, such as balancing inventory costs and spare parts shortage risks while pursuing prediction accuracy, and introducing more refined risk quantification and management methods to build a more comprehensive and practical spare parts management decision support system.

References

1. Zhu, C., Yang, B., Jin, F., Li, M., Jiang, J. (2025). Optimal maintenance and pricing strategy for a periodic review production system with fixed maintenance costs and limited maintenance capacity. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(2). https://doi.org/10.17531/ein/195801
2. Kocer Ozturk, M., Khaniyev, T. (2024). Optimal maintenance policy for a Markov deteriorating system under reliability limit. Eksploatacja i Niezawodność – Maintenance and Reliability, 26(4). https://doi.org/10.17531/ein/190865
3. Su, Z., Hua, Z., Hu, D., Zhao, M. (2025). Research on Equipment Reliability Modeling and Periodic Maintenance Strategies in Dynamic Environment. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(1). https://doi.org/10.17531/ein/192163
4. Ilgin, M. A. (2019). A spare parts criticality evaluation method based on fuzzy AHP and Taguchi loss functions. Eksploatacja i Niezawodność – Maintenance and Reliability, 21(1), 145-152. https://doi.org/10.17531/ein.2019.1.16
5. Yang, D., Guo, H., Wang, Y., & Wang, K. (2025). Research Progress of Lithium-Ion Battery Monitoring Technology Based on Noninvasive Magnetic Induction Sensors. *ACS Applied Electronic Materials*. https://doi.org/10.1021/acsaelm.5c00329
6. Lucht, T., Alieksieiev, V., Kämpfer, T., & et al. (2022). Spare parts demand forecasting in maintenance, repair & overhaul. In *Proceedings of the Conference on Production Systems and Logistics: CPSL 2022* (pp. 525-534). Hannover: publish-Ing.
7. Hemeimat, R., Al-Qatawneh, L., Arafeh, M., & Masoud, S. (2016). Forecasting spare parts demand using statistical analysis. *American Journal of Operations Research*, *6*(02), 113. https://doi.org/10.4236/ajor.2016.62014
8. Yang, Y., Liu, C., & Guo, F. (2017). Forecasting method of aero-material consumption rate based on seasonal ARIMA model. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)* (pp. 2899-2903). IEEE. https://doi.org/10.1109/CompComm.2017.8323062
9. Yang, Y. (2018). Prediction and analysis of aero-material consumption based on multivariate linear regression model. In *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 628-632). IEEE. https://doi.org/10.1109/ICCCBDA.2018.8386591
10. Sahin, M., Kizilaslan, R., & Demirel, Ö. F. (2013). Forecasting aviation spare parts demand using Croston-based methods and artificial neural networks. *Journal of Economic and Social Research, 15*(2), 1.
11. Ferreira, L. M. D. F., Arantes, A., & Silva, C. (2017). Managing service parts for discontinued products: An action research approach. In *International Conference on Operations Research and Enterprise Systems* (pp. 210-223). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-94767-9_11
12. Wang, J., Pan, X., Wang, L., & et al. (2018). Method of spare parts prediction models evaluation based on grey comprehensive correlation degree and association rules mining: A case study in aviation. *Mathematical Problems in Engineering, 2018*(1), 2643405.
13. Hamoud, G. A., & Yiu, C. (2020). Assessment of spare parts for system components using a Markov model. *IEEE Transactions on Power Systems, 35*(4), 3114-3121. https://doi.org/10.1109/TPWRS.2020.2966913
14. Pacheco-Velázquez, E. A., Robles-Cárdenas, M., Juárez Ordóñez, S., & et al. (2023). A heuristic model for spare parts stocking based on

Markov chains. *Mathematics, 11*(16), https://doi.org/10.3390/math11163550

15. Hu, Y. G., Sun, S., & Wen, J. Q. (2014). Agricultural machinery spare parts demand forecast based on BP neural network. *Applied Mechanics and Materials*, *635*, 1822-1825. https://doi.org/10.4028/www.scientific.net/AMM.635-637.1822
16. Jiang, P., Huang, Y., & Liu, X. (2021). Intermittent demand forecasting for spare parts in the heavy-duty vehicle industry: A support vector machine model. *International Journal of Production Research, 59*(24), 7423-7440. https://doi.org/10.1080/00207543.2020.1842936
17. Riwayadi, E., Suprajitno, H., & Miswanto, M. (2024). Enhancing spare parts management through support vector regression: A case study in the service and maintenance industry. *Moneter: Jurnal Keuangan dan Perbankan, 12*(2), 207-217. https://doi.org/10.32832/moneter.v12i2.866
18. Andersson, J., & Siminos, E. (2023). Spare parts demand prediction by using a random forest approach. In *IFIP International Conference on Advances in Production Management Systems* (pp. 793-804). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43670-3_55
19. Chen, F. L., Chen, Y. C., & Kuo, J. Y. (2010). Applying moving back-propagation neural network and moving fuzzy-neuron network to predict the requirement of critical spare parts. *Expert Systems with Applications, 37*(9), 6695-6704. https://doi.org/10.1016/j.eswa.2010.04.037
20. Amin-Naseri, M. R., & Tabar, B. R. (2008). Neural network approach to lumpy demand forecasting for spare parts in process industries. In *2008 International Conference on Computer and Communication Engineering* (pp. 1378-1382). IEEE. https://doi.org/10.1109/ICCCE.2008.4580831
21. Hsu, C. H. (2021). Optimal decision tree for cycle time prediction and allowance determination. *IEEE Access*, 9, 25113-25123. https://doi.org/10.1109/ACCESS.2021.3065391
22. Gomes, M. D., & Garcia, P. A. A. (2023). Spare parts consumption forecasting using a hierarchical Bayesian model. *Pesquisa Operacional, 43*, e269646.
23. Boutselis, P., & McNaught, K. (2019). Using Bayesian Networks to forecast spares demand from Spare Parts failures in a changing service logistics context. *International Journal of Production Economics*, *209*, 325-333. https://doi.org/10.1016/j.ijpe.2018.06.017
24. Sareminia, S. (2023). A support vector-based hybrid forecasting model for chaotic time series: Spare part consumption prediction. *Neural Processing Letters, 55*(3), 2825-2841. https://doi.org/10.1007/s11063-022-10986-4
25. Han, Y., Wang, L., Gao, J., Xing, Z., & Tao, T. (2017, July). Combination forecasting based on SVM and neural network for urban rail vehicle spare parts demand. In *2017 36th chinese control conference (ccc)* (pp. 4660-4665). IEEE. https://doi.org/10.23919/ChiCC.2017.8028090
26. Cui, Z., Jia, H., Gao, Q., & et al. (2024). Maintenance spare parts prediction based on multilevel migration learning CNN-ISE-Attention-BiLSTM. *IEEE Access, 12*, 15208-15221. https://doi.org/10.1109/ACCESS.2024.3357994
27. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003
28. Wang, J., Li, X., Li, J., & et al. (2022). NGCU: A new RNN model for time-series data prediction. *Big Data Research, 27*, 100296.
29. Bi J, Zhang X, Yuan H, et al. A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM[J]. IEEE Transactions on Automation Science and Engineering, 2021, 19(3): 1869-1879. https://doi.org/10.1109/TASE.2021.3077537
30. Reza, S., Ferreira, M. C., Machado, J. J. M., & et al. (2022). A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Systems with Applications, 202*, 117275.
31. Lv, H., Chen, J., Pan, T., Zhang, T., Feng, Y., & Liu, S. (2022). Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application. *Measurement*. https://doi.org/10.1016/j.measurement.2022.111594
32. Zhang X, Guo Y, Shangguan H, et al. Predicting remaining useful life of a machine based on embedded attention parallel networks[J]. Mechanical Systems and Signal Processing, 2023, 192: 110221. https://doi.org/10.1016/j.ymssp.2023.110221
33. Liu, H., Liu, Z., Jia, W., & Lin, X. (2020). Remaining useful life prediction using a novel feature-attention-based end-to-end approach. *IEEE Transactions on Industrial Informatics*. https://doi.org/10.1109/TII.2020.2983760
34. Brauwers, G., & Frasincar, F. (2021). A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*. https://arxiv.org/pdf/2203.14263

35. Feng, Y. W., Chen, J. Y., & Lu, C. (2021). Civil aircraft spare parts prediction and configuration management techniques: review and prospect. *Advances in Mechanical Engineering*. https://journals.sagepub.com/doi/pdf/10.1177/16878140211026173
36. Theodosiou, M. (2011). Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting*, *27*(4), 1178-1195. https://doi.org/10.1016/j.ijforecast.2010.11.002
37. Tian Z W, Qian R L. Chinese Water Demand Forecast Based on iTransformer Model[J]. IEEE Access, 2024. https://doi.org/10.1109/ACCESS.2024.3446663
38. Zhang, D., Zhao, L., Liu, Z., Pan, M., & Li, L. (2025). Temporal-TimesNet: a novel hybrid model for vessel traffic flow multi-step prediction. *Ships and Offshore Structures*. https://www.tandfonline.com/doi/abs/10.1080/17445302.2025.2472438
39. Jensen, S. K., Pedersen, T. B., et al. (2017). Time series management systems: A survey. *IEEE Transactions on Knowledge and Data Engineering*. Retrieved from https://doi.org/10.1109/TKDE.2017.2740932
40. Kantz, H., & Schreiber, T. (2003). Nonlinear time series analysis. *Books.Google.com*. Retrieved from https://books.google.com/books?hl=en&lr=&id=1kUhAwAAQBAJ&oi=fnd&pg=PR11&dq=Spare+parts+consumption+prediction+time+series+decomposition&ots=q0U1aZF2VB&sig=39NmL7CnNsDZmwq63ecc1MIM3Mk
41. Dama, F., & Sinoquet, C. (2021). Time series analysis and modeling to forecast: A survey. *arXiv Preprint*. https://arxiv.org/pdf/2104.00164
42. Wu, H., Liu, Y., Zhou, H., Wang, J., & Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*. https://arxiv.org/pdf/2210.02186
43. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2023). itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.
44. Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., & Wang, Y. (2021). Transformer in transformer. *Advances in neural information processing systems*, *34*, 15908-15919.
45. Zeng, H., Zhang, H., Guo, J., Ren, B., Cui, L., & Wu, J. (2024). A novel hybrid STL-transformer-ARIMA architecture for aviation failure events prediction. *Reliability Engineering & System Safety*, *246*, 110089.
46. Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 15, 5481-5497. Retrieved from https://doi.org/10.5194/gmd-15-5481-2022